

Outage Behavior of Randomly Precoded Integer Forcing Over MIMO Channels

Elad Domanovitz and Uri Erez

Abstract

Integer forcing is an equalization scheme for the multiple-input multiple-output (MIMO) communication channel that has been demonstrated to allow operating close to capacity for “most” MIMO channels. In this work, the measure of “bad” channels is quantified by considering a compound channel setting where the transmitter communicates over a fixed channel but knows only its mutual information. The transmitter encodes the data into independent streams, all taken from the same linear code. The coded streams are then precoded using a unitary matrix. At the receiver side, integer-forcing equalization is applied, followed by standard single-stream decoding. Considering precoding matrices being drawn from a random ensemble, outage corresponds to the event that the target rate exceeds the achievable rate of integer forcing for a given precoding matrix realization. Assuming precoding matrices drawn from the circular unitary ensemble, an explicit universal bound on the outage probability for a given target rate is derived that holds for any channel in the compound class. The derived bound depends only on the gap-to-capacity and number of transmit antennas, and is tight enough to be of interest for moderate values of capacity. The results are also applied to obtain universal bounds on the gap-to-capacity of MIMO closed-loop multicast, achievable via precoded integer forcing.

I. INTRODUCTION

The Multiple-Input Multiple-Output (MIMO) Gaussian channel is central to modern communication and has been extensively studied over the past several decades. Nonetheless, while the capacity limits, under different assumptions on the availability of channel state information, are well understood, the design of low-complexity communication schemes that approach these limits still poses challenges in some scenarios.

For a static channel and a single-user closed-loop setting, capacity may be approached without much difficulty by employing an architecture that decouples coding and modulation. That is, one may use “off-the-shelf” codes in conjunction with linear pre- and post-processing based on matrix decompositions. For instance, one may use the singular-value decomposition (SVD) to transform the channel into parallel scalar additive white Gaussian noise (AWGN) channels, over which standard codes may be employed [1]. Alternatively, standard scalar codes may be used in conjunction with the QR matrix decomposition and successive interference cancellation (SIC) decoding, see, e.g., [2]. Coding for MIMO channels in an ergodic fading environment is more involved but has also been successfully addressed. See, e.g., [3].

In contrast, the focus of this paper is on static (and frequency-flat) MIMO channels where the transmitter only knows (or may only utilize its knowledge of) the mutual information of the channel. More specifically, we address the problem of coding over a compound MIMO channel, as is the case in multicast communication.

Unlike the case of single-user closed-loop communication, designing a transmission scheme based only on the channel’s mutual information (assuming some input covariance matrix, e.g., the scaled identity) is a challenging task as the number of data streams, any linear pre-processing employed, constellation size, and any other transmission design parameters cannot be tailored to the specific channel.

The design of a practical coding scheme for a compound MIMO channel was addressed in [4] where an architecture employing space-time linear processing at the transmitter side and integer-forcing (IF) equalization at the receiver side was proposed. It was shown that such an architecture *universally* achieves the MIMO capacity up to a constant gap, provided the space-time precoding satisfies the non-vanishing determinant (NVD) criterion [5]. The derived gap, however, is very large and thus is of limited practical value.

In the present work, we retain the general architecture of [4], but study its performance when random unitary precoding is applied over the spatial dimension only. Rather than aiming at guaranteeing successful transmission, we study the outage probability of the scheme. Applying random precoding converts the static channel to an effective stochastic one.¹ Further, drawing the precoding matrix from the isotropic unitary ensemble ensures that all channels having the same singular values, will have the same outage probability.

The outage probability in the considered setting thus corresponds to a “scheme outage”. Namely, it is the probability that a random precoding matrix results in an effective channel for which the rate achievable with an IF receiver is smaller than the target rate. In order to provide universal performance guarantees, we study the worst-case outage probability with respect to

The work of E. Domanovitz and U. Erez was supported in part by the Israel Science Foundation under Grant No. 1956/16.

E. Domanovitz and U. Erez are with the Department of Electrical Engineering – Systems, Tel Aviv University, Tel Aviv, Israel (email: domanovi,uri@eng.tau.ac.il).

¹See, e.g., [6] and [7].

all possible singular value combinations corresponding to a given mutual information. Thus, the guaranteed performance does not depend on channel statistics.

We begin by empirically demonstrating that space-only precoded IF (P-IF) has very good performance in terms of worst-case outage probability. We then derive an explicit bound on the latter quantity that depends only on the number of transmit antennas and the gap-to-capacity, where moderate gaps suffice to guarantee a small outage probability.

As another example of an application of the results, we use the probabilistic method to obtain guarantees on the number of users that can be supported in closed-loop MIMO multicast (guaranteeing no outage occurs) as a function of the gap-to-capacity, when using precoded IF.

The paper is organized as follows. Section II defines the channel model of interest and formulates the problem described above. Section III provides background on integer-forcing receivers as well as its use in conjunction with precoding. Section IV derives a universal upper bound for the outage probability of randomly precoded IF over the compound MIMO channel; tighter bounds for the specific case of two transmit antennas for a receiver employing a successive interference cancellation (SIC) variant of IF are also derived. Section V describes the application of the derived bounds to a close-loop MIMO multicast setting.

II. CHANNEL MODEL AND PROBLEM FORMULATION

The single-user (complex) MIMO channel is described by the relation:²

$$\mathbf{y}_c = \mathbf{H}_c \mathbf{x}_c + \mathbf{z}_c, \quad (1)$$

where $\mathbf{x}_c \in \mathbb{C}^{N_t}$ is the channel input vector, $\mathbf{y}_c \in \mathbb{C}^{N_r}$ is the channel output vector, \mathbf{H}_c is an $N_r \times N_t$ complex channel matrix, and \mathbf{z}_c is an additive noise vector of i.i.d. unit variance circularly symmetric complex Gaussian random variables. The input vector \mathbf{x}_c is subject to the power constraint³

$$\mathbb{E}(\mathbf{x}_c^H \mathbf{x}_c) \leq N_t \cdot \text{SNR}. \quad (2)$$

We assume that the channel is fixed throughout the whole transmission period.

The mutual information of the channel (1) is maximized by a Gaussian input [1] with covariance matrix \mathbf{Q}_c satisfying $\text{Tr}(\mathbf{Q}_c) = N_t \text{SNR}$, and is given by

$$C = \max_{\mathbf{Q}_c: \text{Tr}(\mathbf{Q}_c) \leq N_t \text{SNR}} \log \det (\mathbf{I}_{N_r \times N_r} + \mathbf{H}_c \mathbf{Q}_c \mathbf{H}_c^H) \quad (3)$$

For ease of notation, in the sequel we set $\text{SNR} = 1$, i.e., we “absorb” SNR into the channel matrix. Thus, we impose the constraint $\text{Tr}(\mathbf{Q}_c) \leq N_t$ and replace \mathbf{H}_c in (3) $\bar{\mathbf{H}}_c = \mathbf{H}_c \sqrt{\text{SNR}}$ and (with abuse of notation) we omit the bar. The choice of \mathbf{Q}_c that maximizes (3) is determined by the water-filling solution. When the matrix \mathbf{H}_c is known at both transmission ends, i.e., in a closed-loop scenario, this mutual information is the capacity of the channel.

It will prove useful to consider the mutual information achievable isotropic (“white”) input. Specifically, taking $\mathbf{Q}_c = \mathbf{I}_{N_t \times N_t}$, the white-input (WI) mutual information is given by

$$C_{\text{WI}} = \log \det (\mathbf{I} + \mathbf{H}_c \mathbf{H}_c^H). \quad (4)$$

We may define the set

$$\mathbb{H}(C_{\text{WI}}) = \{\mathbf{H}_c \in \mathbb{C}^{N_r \times N_t} : \log \det (\mathbf{I} + \mathbf{H}_c \mathbf{H}_c^H) = C_{\text{WI}}\}, \quad (5)$$

of all channel matrices having the same WI mutual information C_{WI} .

The corresponding compound channel model is defined by (1) with the channel matrix \mathbf{H}_c arbitrarily chosen from the set $\mathbb{H}(C_{\text{WI}})$. The matrix \mathbf{H}_c that was chosen by nature is revealed to the receiver, but not to the transmitter. Clearly, the capacity of this compound channel is C_{WI} , and is achieved with a white Gaussian input.

Employing the IF receiver allows approaching C_{WI} for “most” but not all matrices $\mathbf{H}_c \in \mathbb{H}(C_{\text{WI}})$. We quantify the measure of the set of bad channel matrices by considering outage events, i.e., these events where integer forcing fails even though the channel has sufficient mutual information. More broadly, for a given coding scheme, denote the achievable rate for a given channel matrix \mathbf{H}_c as $R_{\text{scheme}}(\mathbf{H}_c)$. For the case of integer forcing, the explicit expression for $R_{\text{IF}}(\mathbf{H}_c)$ is recalled in Section III-A.

Since applying a precoding matrix \mathbf{P}_c results in an effective channel $\mathbf{H}_c \cdot \mathbf{P}_c$, it follows that the achievable rate of a transmission scheme over this channel is $R_{\text{scheme}}(\mathbf{H}_c \cdot \mathbf{P}_c)$. The worst-case (WC) scheme outage probability is defined in turn

²We denote all complex variables with c to distinguish them from their real-valued representation.

³We denote by $[\cdot]^T$, transpose of a vector and by $[\cdot]^H$, the Hermitian transpose of a vector/matrix.

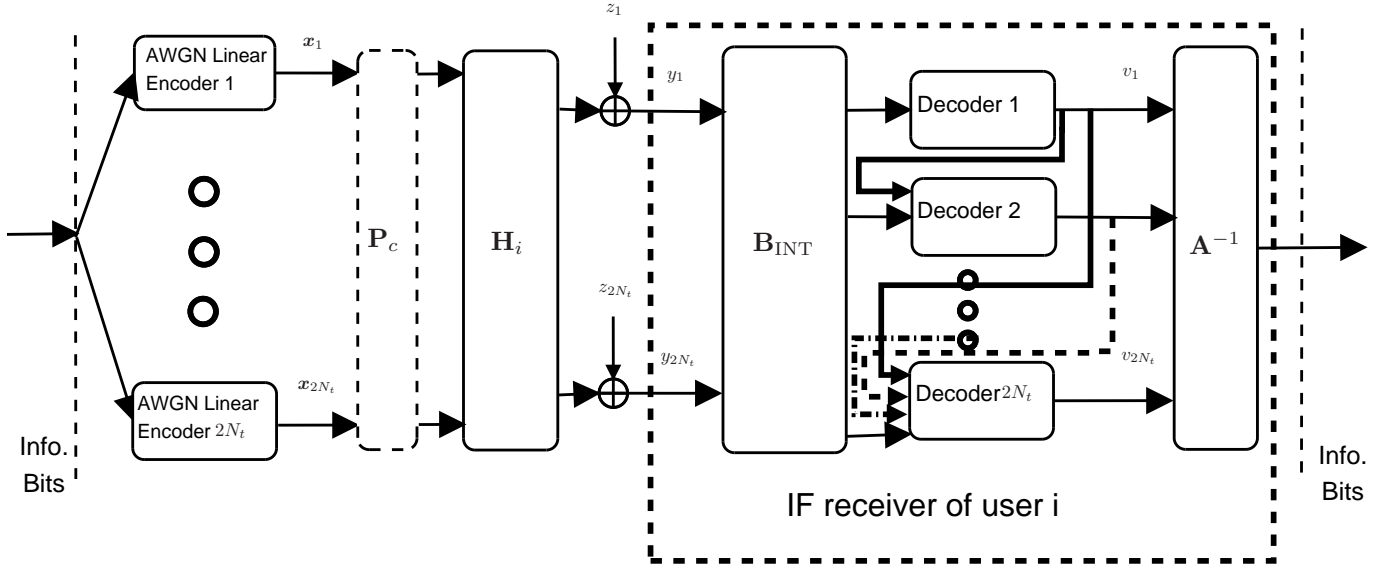


Fig. 1. Precoded-IF-SIC scheme.

as

$$P_{\text{out,scheme}}^{\text{WC}}(C_{\text{WI}}, R) = \sup_{\mathbf{H}_c \in \mathbb{H}(C_{\text{WI}})} \Pr(R_{\text{scheme}}(\mathbf{H}_c \cdot \mathbf{P}_c) < R), \quad (6)$$

where the probability is over the ensemble of precoding matrices. The goal of this paper is to quantify the tradeoff between the transmission rate R and the outage probability $P_{\text{out,IF}}^{\text{WC}}(C_{\text{WI}}, R)$ as defined in (6).

III. INTEGER-FORCING BACKGROUND

A. Single-User Integer-Forcing Equalization

In [8], a receiver architecture scheme coined “integer forcing” was proposed which we briefly recall. We follow the derivation of [8] and describe integer forcing over the reals. Channel model (1) can be expressed via its real-valued representation as

$$\underbrace{\begin{bmatrix} \text{Re}(\mathbf{y}_c) \\ \text{Im}(\mathbf{y}_c) \end{bmatrix}}_{\mathbf{y}} = \underbrace{\begin{bmatrix} \text{Re}(\mathbf{H}_c) & -\text{Im}(\mathbf{H}_c) \\ \text{Im}(\mathbf{H}_c) & \text{Re}(\mathbf{H}_c) \end{bmatrix}}_{\mathbf{H}} \underbrace{\begin{bmatrix} \text{Re}(\mathbf{x}_c) \\ \text{Im}(\mathbf{x}_c) \end{bmatrix}}_{\mathbf{x}} + \underbrace{\begin{bmatrix} \text{Re}(\mathbf{z}_c) \\ \text{Im}(\mathbf{z}_c) \end{bmatrix}}_{\mathbf{z}}.$$

This real-valued representation is used in the sequel to derive performance bounds for the complex channel \mathbf{H}_c . Note that the dimensions of \mathbf{H} are $2N_r \times 2N_t$.

It is assumed that information bits are fed into $2N_t$ encoders, each of which uses the same scalar AWGN *linear* code. The latter produces $2N_t$ channel inputs (for example, x_m for the m ’th antenna).⁴ At the receiver, a *linear equalization* matrix \mathbf{B}_{INT} is applied. It is easiest to understand IF by first describing its zero forcing variant. In this case, \mathbf{B}_{INT} is designed such that the resulting equivalent channel $\mathbf{A} = \mathbf{B}_{\text{INT}}\mathbf{H}$ is a full-rank matrix, all of whose entries are integers. This ensures that the output of the channel (without noise) after applying a modulo operation is a valid codeword. Each of the equalized streams is next passed to a standard (up to the additional element of a modulo operation) AWGN decoder which tries to decode a linear combination of codewords $\mathbf{v}_m = \mathbf{a}_m^T \mathbf{x}$. Finally, after the noise is removed, the original messages are recovered by applying the inverse of \mathbf{A} . Thus, for IF equalization to be successful, decoding over all $2N_t$ subchannels should be correct and the worst subchannel constitutes a bottleneck. The operation of the receiver is depicted in Figure 1 (where at this stage the precoding matrix can be considered as the identity matrix, i.e., $\mathbf{P}_c = \mathbf{I}$).

When using minimum mean square error (MMSE) equalization, rather than zero-forcing, the linear equalizer takes the form

$$\mathbf{B}_{\text{INT}} = \mathbf{A}\mathbf{H}^T (\mathbf{I} + \mathbf{H}\mathbf{H}^T)^{-1}, \quad (7)$$

the input for the m ’th decoder is

$$\mathbf{y}_{\text{eff},m} = \mathbf{v}_m + \mathbf{z}_{\text{eff},m}$$

⁴For simplicity of notation the time index is suppressed as the block length plays no role in our description. Of course, to approach capacity, one needs to use a long block.

where

$$\mathbf{z}_{\text{eff},m} = (\mathbf{b}_m^T \mathbf{H} - \mathbf{a}_m^T) \mathbf{x} + \mathbf{b}_m^T \mathbf{z}.$$

Here, \mathbf{a}_m^T and \mathbf{b}_m^T are the m 'th row of \mathbf{A} and \mathbf{B} respectively. We can define the effective SNR at the m 'th subchannel as

$$\text{SNR}_{\text{eff}}(\mathbf{a}_m) = (\mathbf{a}_m^T (\mathbf{I} + \mathbf{H}^T \mathbf{H})^{-1} \mathbf{a}_m)^{-1},$$

and the effective rate that can be achieved at the m 'th subchannel as

$$R_{\text{IF}}(\mathbf{H}; \mathbf{a}_m) = \frac{1}{2} \log(\text{SNR}_{\text{eff}}(\mathbf{a}_m)) \quad (8)$$

$$= -\frac{1}{2} \log(\mathbf{a}_m^T (\mathbf{I} + \mathbf{H}^T \mathbf{H})^{-1} \mathbf{a}_m). \quad (9)$$

By Theorem 3 in [8], transmission with IF equalizer can achieve any rate satisfying $R < R_{\text{IF}}(\mathbf{H})$ where

$$\begin{aligned} R_{\text{IF}}(\mathbf{H}) &= \max_{\substack{\mathbf{A} \in \mathbb{Z}^{2N_t \times 2N_t} \\ \det \mathbf{A} \neq 0}} \min_{m=1, \dots, 2N_t} 2N_t \cdot R_{\text{IF}}(\mathbf{H}; \mathbf{a}_m) \\ &= \max_{\substack{\mathbf{A} \in \mathbb{Z}^{2N_t \times 2N_t} \\ \det \mathbf{A} \neq 0}} \min_{m=1, \dots, 2N_t} 2N_t \frac{1}{2} \log(\text{SNR}_{\text{eff}}(\mathbf{a}_m)) \\ &= N_t \log \left(\min_{\substack{\mathbf{A} \in \mathbb{Z}^{2N_t \times 2N_t} \\ \det \mathbf{A} \neq 0}} \max_{m=1, \dots, 2N_t} (\mathbf{a}_m^T (\mathbf{I} + \mathbf{H}^T \mathbf{H})^{-1} \mathbf{a}_m)^{-1} \right). \end{aligned} \quad (10)$$

The achievable rates of IF may also be described via the successive minima of a lattice associated with the channel matrix as we now describe. Any channel can be described via its SVD

$$\mathbf{H} = \mathbf{U} \mathbf{\Sigma} \mathbf{V}^T. \quad (11)$$

Using (11), the following decomposition is readily obtained

$$(\mathbf{I} + \mathbf{H}^T \mathbf{H})^{-1} = \mathbf{V} \mathbf{D}^{-1} \mathbf{V}^T, \quad (12)$$

where $\mathbf{D} = \mathbf{I} + \mathbf{\Sigma}^2$. It follows that (9) may be rewritten as

$$R_{\text{IF}}(\mathbf{H}; \mathbf{a}_m) = -\frac{1}{2} \log(\|\mathbf{D}^{-1} \mathbf{V}^T \mathbf{a}_m\|^2) \quad (13)$$

$$\triangleq R_{\text{IF}}^{\mathbf{a}_m}(\mathbf{D}, \mathbf{V}). \quad (14)$$

Let Λ be the lattice spanned by $\mathbf{G} = \mathbf{D}^{-1/2} \mathbf{V}^T$. Recall the definition of successive minima.

Definition 1. Let $\Lambda(\mathbf{G})$ be a lattice spanned by the full-rank matrix $\mathbf{G} \in \mathbb{R}^{K \times K}$. For $k = 1, \dots, K$, we define the k 'th successive minimum as

$$\lambda_k(\mathbf{G}) \triangleq \inf \{r : \dim(\text{span}(\Lambda(\mathbf{G}) \cap \mathcal{B}(\mathbf{0}, r))) \geq k\} \quad (15)$$

where $\mathcal{B}(\mathbf{0}, r) = \{\mathbf{x} \in \mathbb{R}^K : \|\mathbf{x}\| \leq r\}$ is the closed ball of radius r around $\mathbf{0}$. In words, the k 'th successive minimum of a lattice is the minimal radius of a ball centered around $\mathbf{0}$ that contains k linearly independent lattice points.

Thus, the maximal rate achievable with integer forcing equalization (9) may be written as

$$\begin{aligned} R_{\text{IF}}(\mathbf{H}) &= R_{\text{IF}}(\mathbf{D}, \mathbf{V}) \\ &= -2N_t \frac{1}{2} \log(\lambda_{2N_t}^2(\Lambda)) \\ &= N_t \log \left(\frac{1}{\lambda_{2N_t}^2(\Lambda)} \right). \end{aligned} \quad (16)$$

B. Single-User Integer-Forcing Equalization With Successive Interference Cancellation

We also consider a generalized version of the IF mean that incorporates successive interference cancellation. We will refer to it as IF-SIC.⁵ For our purposes, it will suffice to only state the achievable rates of IF-SIC and an operational description of its elements. The reader is referred to [9] for the derivation, details and proofs.

⁵We note that IF-SIC may in general allow using different rates per stream. We nevertheless assume throughout that all streams are encoded via an identical linear code and hence have the same rate.

For a given choice of integer matrix \mathbf{A} , let \mathbf{L} be defined by the following Cholesky decomposition

$$\mathbf{A} (\mathbf{I} + \mathbf{H}^T \mathbf{H})^{-1} \mathbf{A}^T = \mathbf{A} \mathbf{V} \mathbf{D}^{-1} \mathbf{V}^T \mathbf{A}^T \quad (17)$$

$$= \mathbf{L} \mathbf{L}^T. \quad (18)$$

Denoting by $\ell_{m,m}$ the diagonal entries of \mathbf{L} , IF-SIC can achieve (see [9]) any rate satisfying⁶ $R < R_{\text{IF-SIC}}(\mathbf{H})$ where

$$R_{\text{IF-SIC}}(\mathbf{H}) = 2N_t \frac{1}{2} \max_{\mathbf{A}} \min_{m=1, \dots, 2N_t} \log \left(\frac{1}{\ell_{m,m}^2} \right) \quad (19)$$

$$= R_{\text{IF-SIC}}(\mathbf{D}, \mathbf{V}), \quad (20)$$

and the maximization is over all full-rank integer $2N_t \times 2N_t$ matrices. We describe the operation of the IF-SIC receiver, adopting the nomenclature of [9]. First, calculate:

- 1) The optimal integer matrix \mathbf{A} , i.e., the matrix maximizing (20).
- 2) The covariance matrix (17) of the effective noise that arises when using the IF mean \mathbf{B}_{INT} as given in (7).
- 3) The optimal SIC matrix \mathbf{S} as:

$$\mathbf{S} = \text{diag}(\ell_{11}, \dots, \ell_{MM}) \cdot \mathbf{L}^{-1}. \quad (21)$$

- 4) The optimal combined linear front end processing matrix:

$$\begin{aligned} \widetilde{\mathbf{B}}_{\text{INT}} &= \mathbf{S} \mathbf{B}_{\text{INT}} \\ &= \mathbf{S} \mathbf{A} \mathbf{H}^T (\mathbf{I} + \mathbf{H} \mathbf{H}^T)^{-1}. \end{aligned} \quad (22)$$

The operation of the receiver is depicted in Figure 1, where now \mathbf{B}_{INT} is to be understood as $\widetilde{\mathbf{B}}_{\text{INT}}$. Note that this change of linear post-processing is essential to guarantee that the resulting noise variance is minimized. The outputs of decoders $1, \dots, m-1$ are multiplied by $S_{m,1}, \dots, S_{m,m-1}$, respectively, and are then subtracted from the input to decoder m , thereby performing SIC.

C. Precoded Integer-Forcing

1) *Motivating example:* For conventional linear means, bad channels correspond to ill-conditioned matrices. As a concrete example, consider the 2×2 channel

$$\mathbf{H}_{c, \text{WORST}} = \sqrt{2^8 - 1} \begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix}. \quad (23)$$

This channel belongs to the set $\mathbb{H}(C_{\text{WI}} = 8)$ defined in (5). Here, only one transmit and one receive antenna are active so the system reduces to a single-input single-output channel. Thus, the data stream sent from the nonactive antenna is lost. Clearly, in this example, no receiver (including maximum likelihood) will be able to recover the lost data stream and thus the achievable rate of both linear and IF equalization is also zero.

Consider now the channel $\mathbf{H}_{c, \text{WORST}} \cdot \mathbf{P}_c$ where \mathbf{P}_c is a unitary matrix. As the singular values remain unchanged, it is clear that the channel remains ill-conditioned and hence a linear receiver will still suffer from poor performance. On the other hand, IF receivers perform well even over ill-conditioned MIMO channels, and in fact, the performance of the IF receiver for the channel (23) greatly improves for “most” precoding matrices as illustrated below.

2) *Precoding ensemble:* The transmission scheme we analyze consists of applying unitary precoding at the transmitter and IF equalization (either with or without SIC) at the receiver, as depicted in Figure 1. Applying precoding may be viewed as generating a “virtual” channel $\widetilde{\mathbf{H}}_c = \mathbf{H}_c \mathbf{P}_c$ over which transmission takes place.

Throughout this paper, we assume that the precoding matrix \mathbf{P}_c is drawn from what is referred to as the “circular unitary ensemble” (CUE). The ensemble is defined by the unique distribution on unitary matrices that is invariant under left and right unitary transformations [10]. That is, given a random matrix \mathbf{P}_c drawn from the CUE, for any unitary matrix $\check{\mathbf{V}}_c$, both $\mathbf{P}_c \check{\mathbf{V}}_c$ and $\check{\mathbf{V}}_c \mathbf{P}_c$ are equal in distribution to \mathbf{P}_c .

Figure 2 compares the achievable rates of the linear MMSE and IF receivers over the channel (23), when applying a random precoding matrix drawn from the CUE. As can be seen, the achievable rate of IF drastically increases for most precoding matrices, achieving a high fraction of C_{WI} with high probability.

- 3) *Properties of CUE precoding:* The SVD of the precoded channel is

$$\mathbf{H}_c \mathbf{P}_c = \mathbf{U}_c \mathbf{\Sigma}_c \mathbf{V}_c^H \mathbf{P}_c. \quad (24)$$

⁶We note that since we choose to work with equal-rate streams, the constraints on the achievable rate tuples of IF with SIC, as stated in Theorem 2 of [9], play no role in the present work.

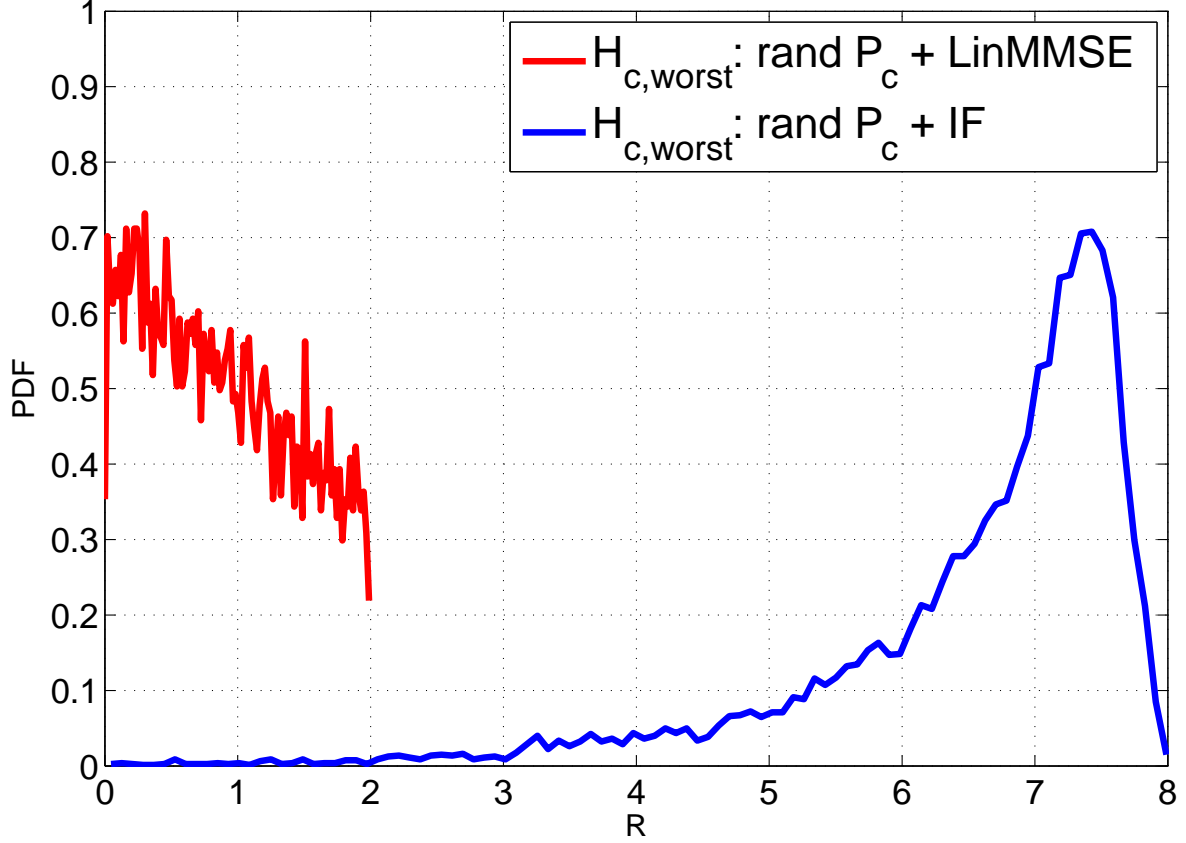


Fig. 2. Approximate probability density function (based on Monte carlo simulation) of achievable rates of the linear MMSE and IF receivers over the channel (23), when applying a random precoding matrix drawn from the CUE.

Since $\mathbf{V}_c^H \mathbf{P}_c$ is equal in distribution to \mathbf{P}_c , for the sake of computing outage probabilities, we may simply assume that \mathbf{V}_c^H (and also \mathbf{V}_c) is drawn from the CUE.

We note that the eigenvalue decomposition of the equivalent real channel can be written as

$$(\mathbf{I} + \mathbf{H}^T \mathbf{H})^{-1} = \mathbf{V} \mathbf{D}^{-1} \mathbf{V}^T, \quad (25)$$

where

$$\mathbf{V} = \begin{bmatrix} \text{Re}(\mathbf{V}_c) & -\text{Im}(\mathbf{V}_c) \\ \text{Im}(\mathbf{V}_c) & \text{Re}(\mathbf{V}_c) \end{bmatrix}. \quad (26)$$

and

$$\mathbf{D} = \begin{bmatrix} \mathbf{D}_c & \mathbf{0} \\ \mathbf{0} & \mathbf{D}_c \end{bmatrix}. \quad (27)$$

Further, the rates of IF, with or without SIC, of such a channel come in pairs.

We may therefore now rewrite (6) as

$$P_{\text{out,IF}}^{\text{WC}}(C, \Delta C) = \sup_{\mathbf{D} \in \mathbb{D}(C)} \Pr(R_{\text{IF}}(\mathbf{D}, \mathbf{V}) < C - \Delta C) \quad (28)$$

where $\mathbb{D}(C)$ is the set of all diagonal matrices \mathbf{D} , with diagonal elements appearing in pairs, such that $\det(\mathbf{D}) = 2^C$.

Another property we use in the sequel is the following. Denote by $d_{c,i}$ the diagonal entries of \mathbf{D}_c . Then

$$2^C = 2^{\log \det(\mathbf{I} + \mathbf{H}_c^H \mathbf{H}_c)} \quad (29)$$

$$= \det(\mathbf{V}_c \mathbf{D}_c \mathbf{V}_c^H) \quad (30)$$

$$= \prod_{i=1}^{N_t} d_{c,i}. \quad (31)$$

Denoting by d_i the diagonal entries of \mathbf{D} , we similarly have

$$2^C = 2^{\frac{1}{2} \log \det(\mathbf{I} + \mathbf{H}^T \mathbf{H})} \quad (32)$$

$$= \sqrt{\det(\mathbf{V} \cdot \mathbf{D} \cdot \mathbf{V}^T)} \quad (33)$$

$$= \prod_{i=1}^{2N_t} \sqrt{d_i}. \quad (34)$$

From (27) we observe that since the singular values of the real channel come in pairs, we have

$$\prod_{i=1}^{2N_t} \sqrt{d_i} = \prod_{i=1}^{N_t} d_{c,i} = 2^C. \quad (35)$$

We denote $d_{\min} = \min_i d_i$ and $d_{\max} = \max_i d_i$.

The following lemma will prove useful in characterizing the performance of precoded IF. It expresses the outage probability of IF for precoding with the CUE in terms of that arising when precoding is performed using the real circular orthogonal ensemble (COE).⁷

Lemma 1. *Let \mathbf{O} be a real $2N_t \times 2N_t$ matrix drawn from the COE. When applying a random complex precoding matrix \mathbf{V}_c which is drawn from the CUE (which induces a real-valued precoding matrix \mathbf{V}), the following holds*

$$\Pr\left(\|\mathbf{D}^{1/2} \mathbf{V} \mathbf{a}\| < \sqrt{\beta}\right) = \Pr\left(\|\mathbf{D}^{1/2} \mathbf{O} \mathbf{a}\| < \sqrt{\beta}\right) \quad (36)$$

Proof: See Appendix A. ■

IV. BOUND ON THE OUTAGE PROBABILITY OF PRECODED INTEGER-FORCING

A. Derivation of Bounds

Define the dual lattice Λ^* which is spanned by the matrix

$$(\mathbf{G}^T)^{-1} = \mathbf{D}^{1/2} \mathbf{V}^T. \quad (37)$$

Recall that the rate of IF is given by (16). Now, the successive minima of Λ and Λ^* are related by (Theorem 2.4 in [11])

$$\lambda_1(\Lambda^*)^2 \lambda_{2N_t}(\Lambda)^2 \leq \frac{2N_t + 3}{4} \gamma_{2N_t}^{*2}, \quad (38)$$

where γ_{2N_t} is Hermite's constant.⁸ Therefore, we may express the achievable rates of IF via the dual lattice as follows

$$R_{\text{IF}}(\mathbf{D}, \mathbf{V}) > N_t \log \left(\frac{\lambda_1^2(\Lambda^*)}{\frac{2N_t+3}{4} \gamma_{2N_t}^{*2}} \right). \quad (39)$$

Hermite's constant is known only for dimensions 1 – 8 and 24. Since it has been never proved that γ_{2N_t} is monotonically increasing, we define

$$\gamma_{2N_t}^* = \max \{\gamma_i : 1 \leq i \leq 2N_t\}. \quad (40)$$

The tightest known bound for Hermite's constant as derived in [13] is

$$\gamma_{2N_t} \leq \left(\frac{2}{\pi}\right) \Gamma(2 + N_t)^{1/N_t}. \quad (41)$$

⁷The COE is defined analogously to the CUE, for the case of real orthonormal matrices [10].

⁸In [12], Theorem 2.1, another bound for the relation between the successive minima of Λ and Λ^* is given. This bound is tighter for very high dimensions (it increases with n^2 , whereas Eq. (38) increases with n^3). However, Eq. (38) has better constants and the cross between these expressions occurs only at $n = 254$.

Since this is an increasing function of N_t , it follows that $\gamma_{2N_t}^*$ is smaller than the r.h.s. of (41).⁹

$$R_{\text{IF}}(\mathbf{D}, \mathbf{V}) \geq N_t \log \left(\frac{\lambda_1^2(\Lambda^*)}{\alpha(N_t)} \right), \quad (42)$$

where

$$\alpha(N_t) = \begin{cases} \frac{2N_t+3}{4} \gamma_{2N_t}^2, & N_t = 2, 3, 4, 12 \\ \frac{2N_t+3}{4} \left(\frac{2}{\pi} \Gamma(2 + N_t)^{1/N_t} \right)^2, & \text{otherwise} \end{cases}. \quad (43)$$

Hence,

$$\Pr(R_{\text{IF}}(\mathbf{D}, \mathbf{V}) < C - \Delta C) \leq \quad (44)$$

$$\Pr \left(N_t \log \left(\frac{\lambda_1^2(\Lambda^*)}{\alpha(N_t)} \right) < C - \Delta C \right) = \Pr \left(\lambda_1^2(\Lambda^*) < 2^{\frac{C-\Delta C}{N_t}} \alpha(N_t) \right). \quad (45)$$

The next lemma gives a bound on the outage probability as a function of the gap-to-capacity ΔC , the capacity C , and d_{\min} (as well as the number of transmit antennas).

Lemma 2. *For any $N_r \times N_t$ complex channel with white-input mutual information C , i.e., $\mathbf{D} \in \mathbb{D}(C)$, and for \mathbf{V}_c drawn from the CUE (inducing a real-valued precoding matrix \mathbf{V}), we have*

$$\Pr(R_{\text{IF}}(\mathbf{D}, \mathbf{V}) < C - \Delta C) \leq \quad (46)$$

$$\sum_{\mathbf{a} \in \mathbb{A}(\beta, d_{\min})} \frac{2N_t \left(2^{\frac{C-\Delta C}{N_t}} \alpha(N_t) \right)^{N_t-1/2}}{\|\mathbf{a}\|^{2N_t-1} 2^C \frac{2}{\sqrt{d_{\min}}}}, \quad (47)$$

where

$$\mathbb{A}(\beta, d_{\min}) = \left\{ \mathbf{a} : 0 < \|\mathbf{a}\| < \sqrt{\frac{\beta}{d_{\min}}} \right\}, \quad (48)$$

with $\beta = 2^{\frac{C-\Delta C}{N_t}} \alpha(N_t)$.

Proof: For some $\beta > 0$, let us upper bound the probability $\Pr(\lambda_1^2(\Lambda^*) < \beta)$ or equivalently $\Pr(\lambda_1(\Lambda^*) < \sqrt{\beta})$. Noting that the event $\{\lambda_1(\Lambda^*) < \sqrt{\beta}\}$ is equivalent to the event

$$\bigcup_{\mathbf{a} \in \mathbb{Z}^{2N_t} \setminus \{0\}} \left\{ \|\mathbf{D}^{1/2} \mathbf{V} \mathbf{a}\| < \sqrt{\beta} \right\} \quad (49)$$

and applying the union bound gives

$$\Pr(\lambda_1(\Lambda^*) < \sqrt{\beta}) < \sum_{\mathbf{a} \in \mathbb{Z}^{2N_t}} \Pr(\|\mathbf{D}^{1/2} \mathbf{V} \mathbf{a}\| < \sqrt{\beta}) \quad (50)$$

$$< \sum_{\mathbf{a} \in \mathbb{A}(\beta, d_{\min})} \Pr(\|\mathbf{D}^{1/2} \mathbf{V}^T \mathbf{a}\| < \sqrt{\beta}), \quad (51)$$

where the second inequality follows since whenever $\|\mathbf{a}\| \cdot \sqrt{d_{\min}} \geq \sqrt{\beta}$, we have that $\Pr(\|\mathbf{D}^{1/2} \mathbf{V} \mathbf{a}\| < \sqrt{\beta}) = 0$.

Let \mathcal{S} denote the unit sphere of dimension $2N_t$, i.e.,

$$\mathcal{S} = \{(x_1, x_2, \dots, x_{2N_t}) : x_1^2 + x_2^2 + \dots + x_{2N_t}^2 = 1\}. \quad (52)$$

By Lemma 1

$$\Pr(\|\mathbf{D}^{1/2} \mathbf{V} \mathbf{a}\| < \sqrt{\beta}) = \Pr(\|\mathbf{D}^{1/2} \mathbf{O} \mathbf{a}\| < \sqrt{\beta}). \quad (53)$$

Let $\mathbf{o}_{\|\mathbf{a}\|} \sim \text{Unif}(\mathcal{S} \cdot \|\mathbf{a}\|)$, and note that $\mathbf{O} \mathbf{a}$ is equal in distribution to $\mathbf{o}_{\|\mathbf{a}\|}$. It follows that

$$\Pr(\|\mathbf{D}^{1/2} \mathbf{V} \mathbf{a}\| < \sqrt{\beta}) = \Pr(\|\mathbf{D}^{1/2} \mathbf{o}_{\|\mathbf{a}\|}\| < \sqrt{\beta}). \quad (54)$$

⁹In the sequel we use known values of Hermite's constant when possible, i.e. for $N_t = 2, 3, 4$. For other dimensions, we use this bound.

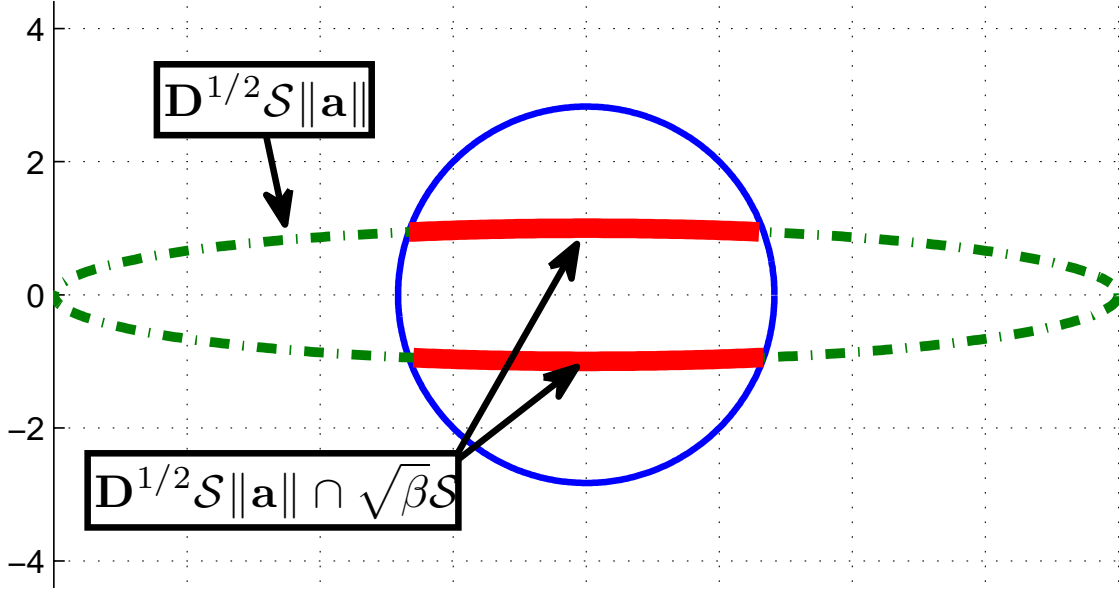


Fig. 3. Illustration of the geometric objects appearing in (55-59).

Now the probability appearing on the r.h.s. of (54) has a simple geometric interpretation. Define an ellipsoid with axes $x_i = \sqrt{d_i} \cdot \|\mathbf{a}\|$ and denote its surface area by $L(x_1, x_2, \dots, x_{2N_t})$. Then, the r.h.s. of (54) is the ratio of the part of the entire surface area of an ellipsoid which is inside a sphere of radius $\sqrt{\beta}$ (denoted by $\text{CAP}_{\text{ell}}(x_1, x_2, \dots, x_{2N_t})$) and the total surface area of the ellipsoid. This is illustrated in Figure 3 for the case of two real dimensions. We may rewrite (54) as

$$\Pr\left(\|\mathbf{D}^{1/2}\mathbf{o}_{\|\mathbf{a}\|}\| < \sqrt{\beta}\right) = \frac{|\mathbf{D}^{1/2}\mathcal{S} \cdot \|\mathbf{a}\| \cap \sqrt{\beta}\mathcal{S}|}{|\mathbf{D}^{1/2}\|\mathbf{a}\|\mathcal{S}|} \quad (55)$$

$$= \frac{\text{CAP}_{\text{ell}}(x_1, x_2, \dots, x_{2N_t})}{L(x_1, x_2, \dots, x_{2N_t})}. \quad (56)$$

Neither the numerator nor the denominator of (56) has a closed form expression. In order to upper bound this ratio, we upper bound the numerator $\text{CAP}_{\text{ell}}(x_1, x_2, \dots, x_{2N_t})$ and lower bound the denominator (the surface area of the ellipsoid).

Using (54) and (57) in [14], we have

$$L(x_1, x_2, \dots, x_{2N_t}) > \Omega_{2N_t} \|\mathbf{a}\|^{2N_t} \prod_{i=1}^{2N_t} \sqrt{d_i} \sum_{i=1}^{2N_t} \frac{1}{\|\mathbf{a}\| \sqrt{d_i}} \quad (57)$$

$$> \Omega_{2N_t} \|\mathbf{a}\|^{2N_t-1} 2^C \frac{2}{\sqrt{d_{\min}}} \quad (58)$$

$$\triangleq \underline{L}(x_1, x_2, \dots, x_{2N_t}), \quad (59)$$

where $\Omega_{2N_t} = \frac{\pi^{N_t}}{\Gamma(1+N_t)}$ is the volume of a unit ball of dimension $2N_t$.

As an upper bound for the numerator, we take the entire surface area of a sphere of radius $\sqrt{\beta}$, which is given by

$$A_{2N_t}(\sqrt{\beta}) = 2N_t \frac{\pi^{N_t}}{\Gamma(1+N_t)} \sqrt{\beta}^{2N_t-1}. \quad (60)$$

We thus have

$$\text{CAP}_{\text{ell}}(x_1, x_2, \dots, x_{2N_t}) \leq A_{2N_t}(\sqrt{\beta}) \quad (61)$$

$$\triangleq \overline{\text{CAP}_{\text{ell}}}(\sqrt{\beta}). \quad (62)$$

We may therefore bound (56) by

$$\Pr\left(\|\mathbf{D}^{1/2}\mathbf{o}_{\|\mathbf{a}\|}\| < \sqrt{\beta}\right) \leq \frac{\overline{\text{CAP}_{\text{ell}}}(\sqrt{\beta})}{\underline{L}(x_1, x_2, \dots, x_{2N_t})}. \quad (63)$$

Substituting (59), (62) into (63) yields

$$\sum_{\mathbb{A}(\beta, d_{\min})} \Pr \left(\|\mathbf{D}^{1/2} \mathbf{o}_{\|\mathbf{a}\|}\| < \sqrt{\beta} \right) < \quad (64)$$

$$\sum_{\mathbb{A}(\beta, d_{\min})} \frac{2N_t \frac{\pi^{N_t}}{\Gamma(1+N_t)} \sqrt{\beta}^{2N_t-1}}{\frac{\pi^{N_t}}{\Gamma(1+N_t)} \|\mathbf{a}\|^{2N_t-1} 2^C \frac{2}{\sqrt{d_{\min}}}} \quad (65)$$

$$= \sum_{\mathbb{A}(\beta, d_{\min})} \frac{2N_t \sqrt{\beta}^{2N_t-1}}{\|\mathbf{a}\|^{2N_t-1} 2^C \frac{2}{\sqrt{d_{\min}}}}. \quad (66)$$

Substituting $\beta = 2^{\frac{C-\Delta C}{N_t}} \cdot \alpha(N_t)$, we finally arrive at

$$\Pr(R_{\text{IF}}(\mathbf{D}, \mathbf{V}) < C - \Delta C) \quad (67)$$

$$\leq \sum_{\mathbb{A}(\beta, d_{\min})} \Pr \left(\|\mathbf{D}^{1/2} \mathbf{o}_{\|\mathbf{a}\|}\| < \sqrt{\beta} \right) \quad (68)$$

$$\leq \sum_{\mathbb{A}(\beta, d_{\min})} \frac{2N_t \left(2^{\frac{C-\Delta C}{N_t}} \alpha(N_t) \right)^{N_t-1/2}}{\|\mathbf{a}\|^{2N_t-1} 2^C \frac{2}{\sqrt{d_{\min}}}}. \quad (69)$$

■

The bound of Lemma 2 is depicted in Figure 4. Rather than plotting the outage probability, its complement is depicted, i.e., we depict the cumulative distribution function of the achievable rate. For given C and ΔC , Lemma 2 was numerically calculated over a grid of singular values. For each such vector of singular values, calculation was performed over all $\mathbf{a} \in \mathbb{A}(\beta, d_{\min})$. The worst-case outage probability over all vectors of singular values from the grid is presented. The integer matrix was derived using the LLL algorithm. Advanced techniques exist (see, e.g., [15] and [16]) which can be used to further improve the numerical results.

In addition, empirical (Monte Carlo) results are also plotted. For each vector of singular values, a large number of random unitary matrices was drawn and the outage probability was calculated. Then, the worst case outage probability over all tested (i.e., those belonging to the grid) singular values is presented.

As a further reference, the figure also depicts the universal guaranteed gap-to-capacity derived in [4], which for the case of $N_t = 2$, amounts to $\Delta C = 15.24$ bits [4].¹⁰

While Lemma 2 provides an explicit bound on the outage probability, in order to calculate it, one needs to go over all diagonal matrices in $\mathbb{D}(C)$ and for each diagonal matrix, sum over all the relevant integer vectors in $\mathbb{A}(\beta, d_{\min})$. Hence, the bound can be evaluated only for moderate capacities and a small number of transmit antennas. The following theorem provides (a looser) simple closed-form bound. Furthermore, this bound does not depend on capacity but rather only on the number of transmit antennas and the gap-to-capacity.

Theorem 1. *For any $N_r \times N_t$ complex channel with WI mutual information C , and for \mathbf{V}_c drawn from the CUE (which induces a real-valued precoding matrix \mathbf{V}), we have*

$$P_{\text{out,IF}}^{\text{WC}}(C, \Delta C) \leq c(N_t) 2^{-\Delta C}, \quad (70)$$

where

$$c(N_t) = \left[\left(2 + \frac{\sqrt{2N_t}}{2} \right)^{2N_t} - \left(1 - \frac{\sqrt{2N_t}}{2} \right)^{2N_t} \right] N_t \alpha(N_t)^{N_t} \frac{\pi^{N_t}}{\Gamma(N_t+1)} \quad (71)$$

and

$$\alpha(N_t) = \frac{2N_t + 3}{4} \left(\frac{2}{\pi} \Gamma(2 + N_t)^{1/N_t} \right)^2. \quad (72)$$

Thus, $c(N_t)$ is a constant that depends only on N_t .

Proof: See Appendix B. ■

This bound is also depicted in Figure 4 for the case of two transmit antennas. The bound is depicted in Figure 5 (solid lines) for other values of N_t . Recall again that, for $N_t = 2, 3, 4$, we use the actual values of $\gamma_4 = \sqrt{2}$, $\gamma_6 = \left(\frac{64}{3}\right)^{1/6}$ and $\gamma_8 = 2$, rather than the bound of [13].

¹⁰This upper bound on the gap-to-capacity is guaranteed for a different coding scheme than that considered in this paper, combining NVD precoding with space-time modulation. Nevertheless, it serves as a useful benchmark.

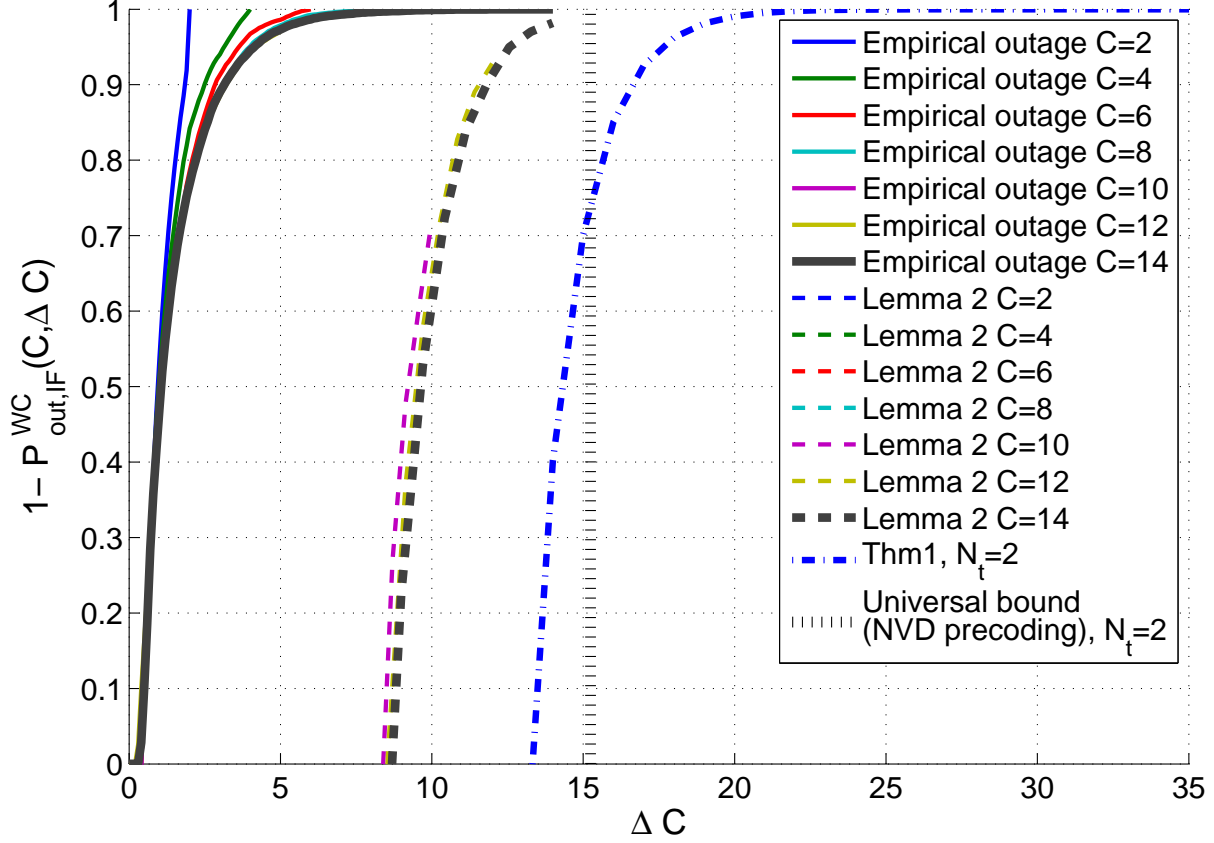


Fig. 4. Comparing empirical results, Lemma 2 and Theorem 1 for $N_r \times 2$ complex channels for various values of capacity.

B. Asymptotic Behaviour of the Bounds for Rayleigh Fading

It is of interest to analyze the high SNR asymptotics of the derived bounds for a statistical channel model. The diversity-multiplexing tradeoff (DMT) provides a rough characterization of the performance of a MIMO transmission scheme at high SNR [17]. Recall the standard definition for multiplexing gain r and diversity gain d of a scheme [17]:

$$\lim_{\text{SNR} \rightarrow \infty} \frac{R(\text{SNR})}{\log \text{SNR}} = r \quad (73)$$

$$\lim_{\text{SNR} \rightarrow \infty} -\frac{\log p_{\text{error}}(\text{SNR})}{\text{SNR}} = d, \quad (74)$$

where $R(\text{SNR})$ and $p_{\text{error}}(\text{SNR})$ are the rate and probability of error of the scheme at a given SNR value.¹¹

For i.i.d. Rayleigh fading, and assuming $N_r \geq N_t$, it was shown in [8] that, IF achieves the optimal receive DMT, i.e., that

$$d_{\text{IF}}(r) = N_r \left(1 - \frac{r}{N_t} \right). \quad (75)$$

We note that i.i.d. Rayleigh fading is invariant to CUE precoding and thus the optimal receive DMT is also achieved when applying the random precoding considered in this paper.

The following lemma gives the DMT that may be established using the bound for the worst-case outage probability of Theorem 1.

Lemma 3. *The achievable DMT based on the worst-case bound of Theorem 1 is*

$$d_{\text{Thm1}}(r) = \min(N_t, N_r) - r. \quad (76)$$

¹¹Note that $p_{\text{error}}(\text{SNR})$ now represents the total probability of error, whether it is a result of the channel being in outage or is the outcome of a scheme outage.

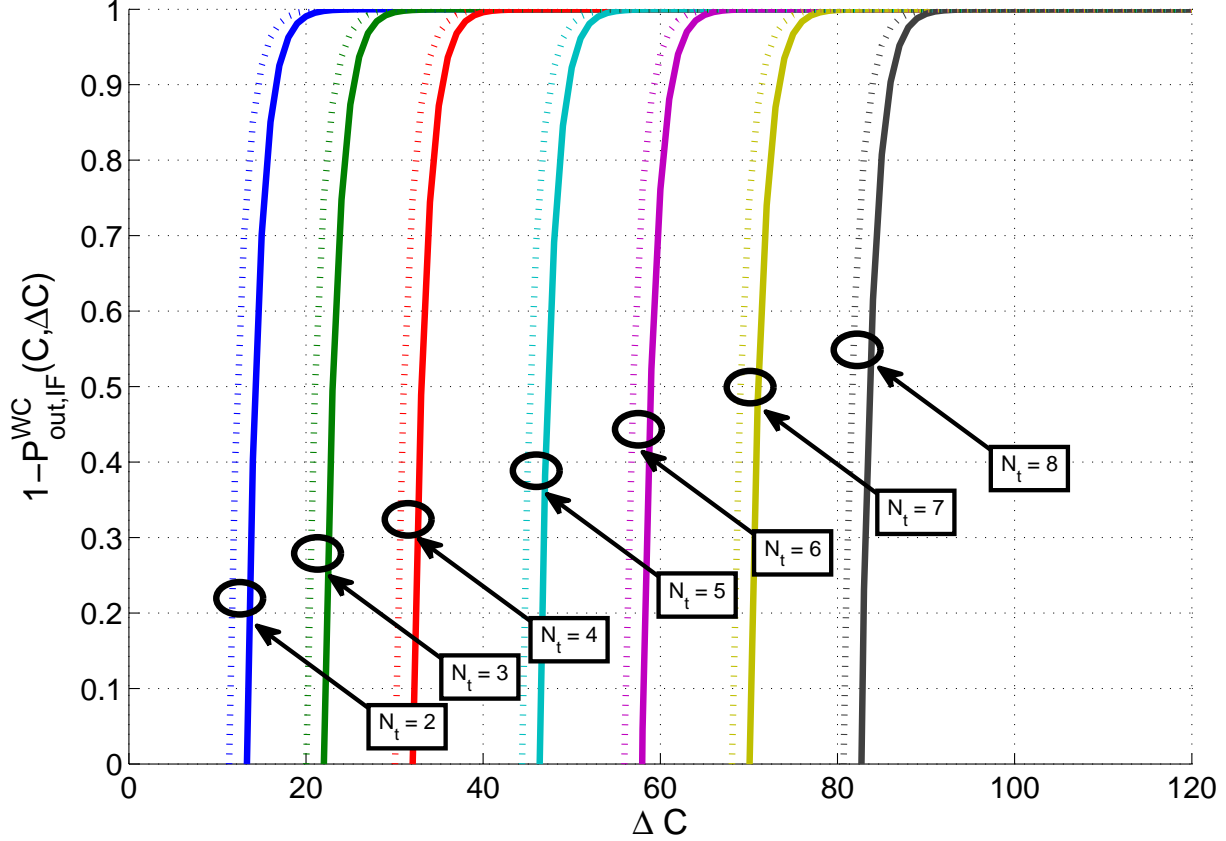


Fig. 5. The performance guaranteed by Theorem 1 for $N_t = 2, 3, \dots, 8$ transmit antennas.

Proof: See Appendix D. ■

The DMT of Lemma 3 coincides with the optimal receive DMT (which is indeed attained by the IF receiver assuming $N_r \geq N_t$, as shown in [8]) only for the case where $N_t = N_r$. In all other cases, it is inferior to the true DMT. We believe that this is unavoidable when deriving bounds based on the worst-case member in the compound channel class.

C. Improving the Bounds

A close inspection of the bound of Theorem 1 reveals that there are two main sources for looseness that may be further tightened:

- *Union bound* - While there is an inherent loss in the union bound, some terms in the summation (51) may be completely dropped. See Corollary 1 below.
- *Dual Lattice* - Bounding via the dual lattice induces a loss reflected in (38). This may be circumvented for the case of $N_t = 2$ transmit antennas, as accomplished (along with other improvements) in Theorem 2 below.

We first tighten the union bound. As expressed in (49), the event where the first of the successive minima is smaller than $\sqrt{\beta}$ is equivalent to going over all integer vectors and checking whether any of them meet the norm condition. However, going over *all* integer vectors is superfluous. In case that an integer vector $\mathbf{b} \in \mathbb{A}(\beta, \mathbf{d}_{\min})$ is an integer multiple of another integer vector $\mathbf{a} \in \mathbb{A}(\beta, \mathbf{d}_{\min})$, there is no need to check both of them. Rather, it suffices to check and include in the union bound only the event corresponding to \mathbf{a} .

It follows that one may replace the set $\mathbb{A}(\beta, \mathbf{d}_{\min})$ appearing in the summation in (2) by a smaller set $\mathbb{B}(\beta, \mathbf{d}_{\min})$ as described by the next corollary.

Corollary 1. For any $N_r \times N_t$ complex channel and for \mathbf{V} drawn from the CUE, we have

$$\Pr(R_{\text{IF}}(\mathbf{D}, \mathbf{V}) < C - \Delta C) \leq \quad (77)$$

$$\sum_{\mathbf{a} \in \mathbb{B}(\beta, d_{\min})} \frac{2N_t \left(2^{\frac{C-\Delta C}{N_t}} \alpha(N_t) \right)^{N_t-1/2}}{\|\mathbf{a}\|^{2N_t-1} 2^C \frac{2}{\sqrt{d_{\min}}}}. \quad (78)$$

where

$$\mathbb{B}(\beta, d_{\min}) = \quad (79)$$

$$\left\{ \mathbf{a} : 0 < \|\mathbf{a}\| < \sqrt{\frac{\beta}{d_{\min}}} \text{ and } \nexists 0 < c < 1 \text{ s.t. } c\mathbf{a} \in \mathbb{Z}^n \right\}, \quad (80)$$

with $\beta = 2^{\frac{C-\Delta C}{N_t}} \alpha(N_t)$.

Another improvement can be obtained by noting that \mathbf{D} and \mathbf{V} are the real representations of complex matrices. Similarly, the integer vector \mathbf{a} may be viewed as a real representation of the complex vector \mathbf{a}_c . As multiplication of \mathbf{a}_c by $\{-1, j, -j\}$ does not change the value of $\|\mathbf{D}^{1/2} \mathbf{V} \mathbf{a}\|$, it suffices to include only one of these members of $\mathbb{A}(\beta, d_{\min})$ in the summation. Hence, a simple multiplicative improvement may be obtained.

Corollary 2.

$$\Pr(R_{\text{IF}}(\mathbf{D}, \mathbf{V}) < C - \Delta C) \leq \quad (81)$$

$$\frac{1}{4} \sum_{\mathbf{a} \in \mathbb{A}(\beta, d_{\min})} \frac{2N_t \beta^{N_t-1/2}}{\|\mathbf{a}\|^{2N_t-1} 2^C \frac{2}{\sqrt{d_{\min}}}}, \quad (82)$$

with $\beta = 2^{\frac{C-\Delta C}{N_t}} \alpha(N_t)$.

While the improvement of Corollary 1 depends on β (and hence also on C), we may tighten Theorem 1 by invoking Corollary 2 as shown by the dashed lines in Figure 5. For a given value of C , we may combine the two corollaries. Figure 6 shows the different bounds on the outage probability for the case of $N_r \times 2$ MIMO channel with $C = 14$, where both corollaries are utilized for tightening Lemma 2.

As mentioned above, there is an additional significant loss due to using the dual lattice for deriving both Lemma 2 and Theorem 1. For the case of $N_t = 2$, this loss can be circumvented by analyzing the performance of IF-SIC. When using IF-SIC, (6) can be rewritten as

$$P_{\text{out,IF-SIC}}^{\text{WC}}(C, \Delta C) = \sup_{\mathbf{D} \in \mathbb{D}(C)} \Pr(R_{\text{IF-SIC}}(\mathbf{D}, \mathbf{V}) < C - \Delta C). \quad (83)$$

The next lemma provides a bound on the outage probability of IF-SIC.

Lemma 4. For any $N_r \times 2$ complex channel with white-input mutual information $C > 1$, i.e., $\mathbf{D} \in \mathbb{D}(C)$, and for \mathbf{V}_c drawn from the CUE (which induces a real-valued precoding matrix \mathbf{V}), we have

$$\Pr(R_{\text{IF-SIC}}(\mathbf{D}, \mathbf{V}) < C - \Delta C) < \quad (84)$$

$$\sum_{\mathbf{a} \in \mathbb{A}(\beta, d_{\max})} \frac{2\pi^2 2^{3/4(C+\Delta C)}}{\pi^2 \frac{\|\mathbf{a}\|^3}{2^C} (\sqrt{d_{\max}})}, \quad (85)$$

where

$$\mathbb{A}(\beta, d_{\max}) = \left\{ \mathbf{a} : 0 < \|\mathbf{a}\| < \sqrt{\beta d_{\max}} \right\}. \quad (86)$$

with $\beta = 2^{1/2(C-\Delta C)}$.

Proof: See Appendix C. ■

Remark 1. Corollary 1 holds also for the case of $N_t = 2$. Hence, Lemma 4 can be further tightened by replacing $\mathbb{A}(\beta, d_{\min})$ with $\mathbb{B}(\beta, d_{\min})$.

In a similar manner to the derivation of Theorem 1 using Lemma 2, for IF-SIC, Lemma 4 leads to the following theorem.

Theorem 2. For any $N_r \times 2$ complex channel \mathbf{H}_c with white-input mutual information $C > 1$, i.e., $\mathbf{D} \in \mathbb{D}(C)$, and for \mathbf{V}_c

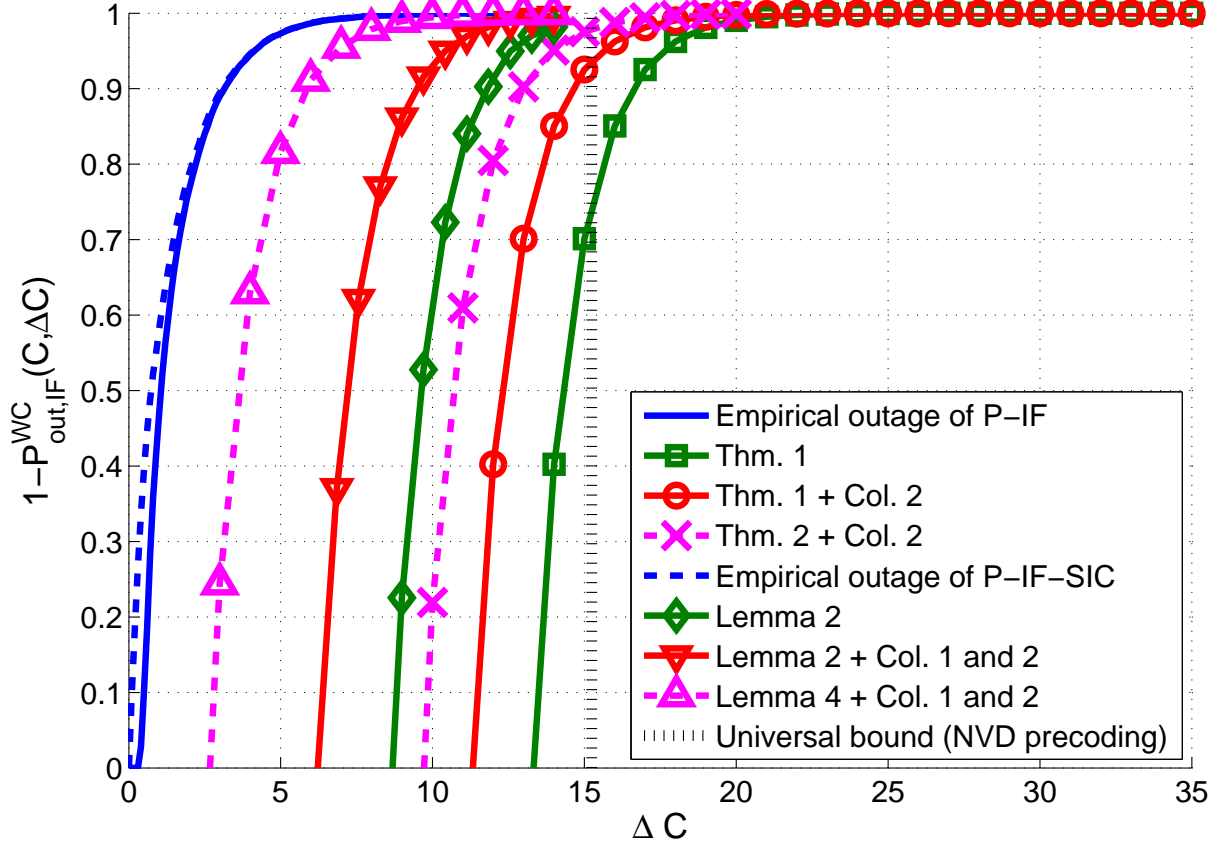


Fig. 6. Outage bounds for $N_r \times 2$ channel while $C = 14$ bits.

drawn from the CUE (which induces a real-valued precoding matrix \mathbf{V}), we have

$$P_{\text{out,IF-SIC}}^{\text{WC}}(C, \Delta C) \leq 81\pi^2 2^{-\Delta C}, \quad (87)$$

for all $\Delta C > 1$ (i.e., for $R_{\text{IF-SIC}}(\mathbf{D}, \mathbf{V}) < C - 1$).

Proof: See Appendix C. ■

Figure 6 depicts the improved bounds of Lemma 4 (incorporating the improvement provided by Corollary 1 and 2) and Theorem 2 for a system employing IF-SIC with $N_t = 2$.

D. Comparison with maximum-likelihood decoding

Beyond the performance bounds derived thus far, it is natural to compare the performance attained by an IF receiver with that of an optimal maximum likelihood (ML) decoder for the same randomly precoded scheme but where each stream is coded using an independent Gaussian codebooks.

Let \mathbf{H}_S denote the submatrix of \mathbf{H}_c formed by taking the columns with indices in $S \subseteq 1, 2, \dots, 2M$. If we use a joint ML detector that searches for the most likely set of transmitted messages then the following rate is achievable:

$$R_{\text{JOINT}} = \min_{S \subseteq 1, 2, \dots, N_t} \frac{N_t}{|S|} \log \det (\mathbf{I}_S + \mathbf{H}_S \mathbf{H}_S^H). \quad (88)$$

Figure 7 provides a comparison between the empirical performance of P-IF-SIC and the performance of the corresponding scheme with ML decoding. As can be seen, for small outage probabilities the gap is quite small and grows up to 0.5 bit for high outage probabilities. This suggests that most of the loss with respect to the WI mutual information is due to the separate coding of the data streams and not due to the suboptimality of the IF receiver.

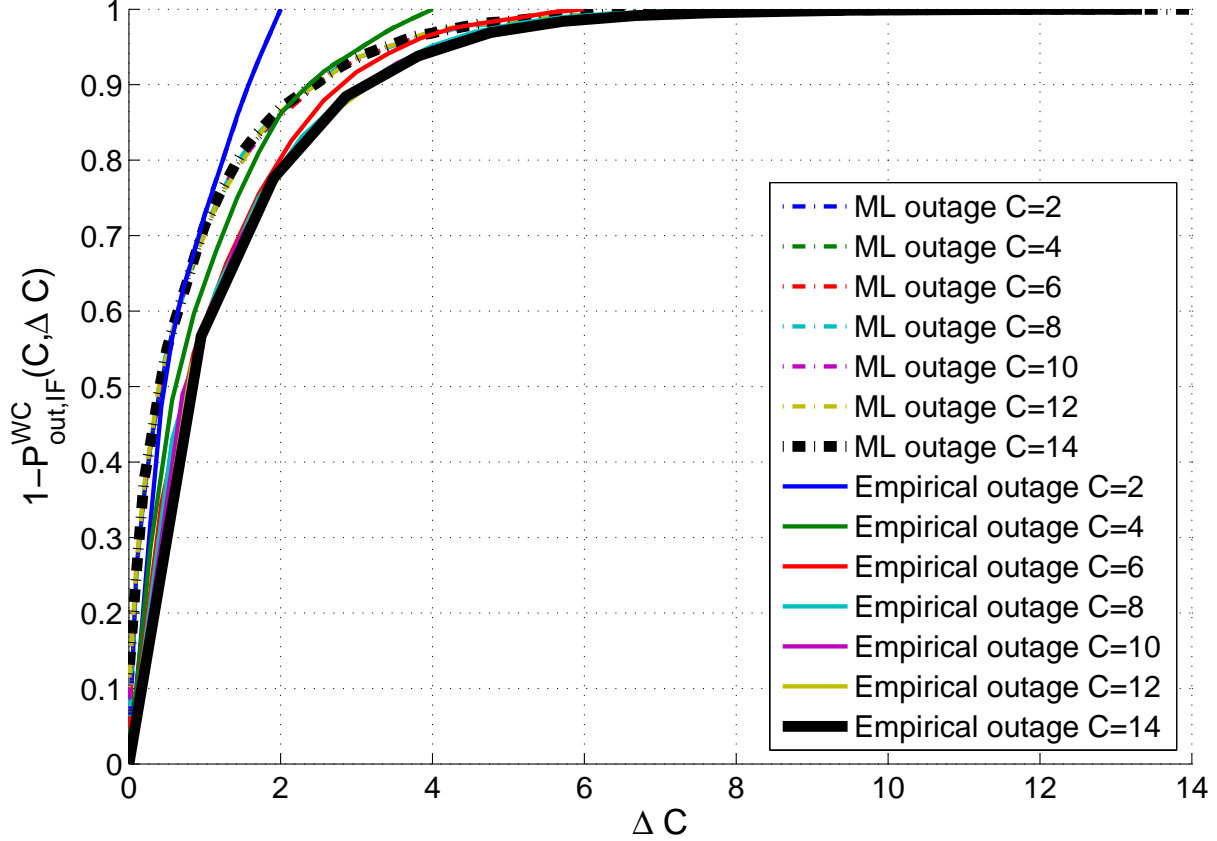


Fig. 7. Performance comparison between P-IF-SIC and joint ML decoding.

V. APPLICATION : UNIVERSAL GAP-TO-CAPACITY FOR MULTI-USER CLOSED-LOOP MULTICAST USING P-IF

Closed-loop MIMO multicast is a scenario where a transmitter equipped with N_t transmit antennas wishes to send the same message to K users, where user i is equipped with N_i antennas.

Even though channel state information is available at both transmission ends, designing practical capacity-approaching schemes for closed-loop multicast with $K \geq 3$ users is challenging as detailed in [18]. Specifically, to achieve a small gap to capacity, the scheme of [18] requires utilizing space-time coding with a large number of channel uses. The outage bound derived above suggests that P-IF may be an attractive practical closed-loop MIMO multicast scheme, allowing to obtain a small gap to capacity with space-only precoding.

Denoting by $\mathbf{H}_{c,i}$ the $N_i \times N_t$ channel matrix corresponding to the i th user (and by $\mathcal{H} = \{\mathbf{H}_{c,i}\}_{i=1}^K$ the set of channels), the received signal at user i is

$$\mathbf{y}_{c,i} = \mathbf{H}_{c,i} \mathbf{x}_c + \mathbf{z}_{c,i}. \quad (89)$$

We assume that channel state information (CSI) is available at both transmission ends.

The multicast capacity is defined as the capacity of the compound channel (89). It is attained by a Gaussian vector input, where the mutual information is maximized over all covariance matrices \mathbf{Q}_c satisfying $\text{Tr}(\mathbf{Q}_c) \leq N_t$:

$$C(\mathcal{H}) = \max_{\mathbf{Q}_c: \text{Tr}(\mathbf{Q}_c) \leq N_t} \min_{\mathbf{H}_c \in \mathcal{H}} \log \det(\mathbf{I} + \mathbf{H}_c \mathbf{Q}_c \mathbf{H}_c^H). \quad (90)$$

We assume without loss of generality that the input covariance matrix is the identity matrix. We may do so since the covariance shaping matrix $\mathbf{Q}_c^{1/2}$ may be absorbed into the channel by defining the effective channel $\hat{\mathbf{H}}_{c,i} = \mathbf{H}_{c,i} \mathbf{Q}_c^{1/2}$. Thus,

$$C(\mathcal{H}) = \min_i C_{\text{WI}}(\hat{\mathbf{H}}_{c,i}). \quad (91)$$

In other words, after finding the optimal covariance matrix \mathbf{Q}_c , when it comes to the transmission scheme, it suffices to consider

WI transmission over the effective channels $\hat{\mathbf{H}}_{c,i}$. With a slight abuse of notation we use $\mathbf{H}_{c,i}$ to denote the effective channel, i.e., we drop the hat. We note that for each user i , there exists an $\alpha_i \geq 1$ such that

$$\mathbf{H}_{c,i} = \alpha_i \check{\mathbf{H}}_{c,i}. \quad (92)$$

where

$$\check{\mathcal{H}} = \{\check{\mathbf{H}}\}_{i=1}^K \in \mathbb{H}(C(\mathcal{H})), \quad (93)$$

i.e., $\{\check{\mathbf{H}}\}_{i=1}^K$ is contained in the (continuum) set of channels, having the same capacity $C(\mathcal{H})$. Further, α_i can be interpreted as excess SNR that user i has, beyond the minimum it needs in the multicast setting. Since the achievable rate of IF is monotonically increasing in SNR, it follows that the achievable rates over the set of channels \mathcal{H} can only be higher than over $\check{\mathcal{H}}$, which we next lower bound.

Let us consider applying the random precoded scheme of Section IV to the compound channel set $\check{\mathcal{H}}$.¹² Define $A_i(R)$ as the event where a random precoding matrix \mathbf{P}_c drawn from a CUE achieves a desired target R for user i

$$A_i(R) = \{\mathbf{P}_c : R_{\text{IF}}(\mathbf{H}_{c,i} \cdot \mathbf{P}_c) \geq R\}. \quad (94)$$

We are interested in the probability of achieving the target rate for all users, i.e., $\Pr(\cap A_i(R))$. Note that

$$\Pr(\cap A_i(R)) = 1 - \Pr(\overline{\cap A_i(R)}) = 1 - \Pr(\cup \overline{A_i(R)}). \quad (95)$$

Applying the union bound, we get

$$\Pr(\cup \overline{A_i(R)}) \leq \sum \Pr(\overline{A_i(R)}). \quad (96)$$

Thus,

$$\Pr(\cap A_i(R)) \geq 1 - \sum \Pr(\overline{A_i(R)}). \quad (97)$$

Define

$$\check{A}_i(R) = \{\mathbf{P}_c : R_{\text{IF-SIC}}(\check{\mathbf{H}}_{c,i} \cdot \mathbf{P}_c) \geq R\}. \quad (98)$$

Recalling (93), and noticing that $\Pr(A_i(R))$ is defined as the probability of achieving the target rate, whereas $P_{\text{out,IF}}^{\text{WC}}$ is defined as the probability of failing to achieve the target rate, we have

$$\Pr(A_i(R)) \geq \Pr(\check{A}_i(R)) \geq 1 - P_{\text{out,IF}}^{\text{WC}}(C(\mathcal{H}), C - R), \quad (99)$$

or equivalently,

$$\Pr(\overline{A_i(R)}) = 1 - \Pr(A_i(R)) \leq P_{\text{out,IF}}^{\text{WC}}(C(\mathcal{H}), C - R). \quad (100)$$

It follows that,

$$\Pr(\cap A_i(R)) \geq 1 - K P_{\text{out,IF}}^{\text{WC}}(C(\mathcal{H}), C - R). \quad (101)$$

This provides a means to obtain a guaranteed achievable transmission rate $R_{\text{WC-CL}}(\mathcal{H})$ for closed-loop precoded IF. Namely, $R_{\text{WC-CL}}(\mathcal{H})$ is the maximum rate for which

$$P_{\text{out,IF}}^{\text{WC}}(C(\mathcal{H}), C - R) \leq \frac{1}{K}. \quad (102)$$

Substituting (102) in (101) we get that any $R < R_{\text{WC-CL}}(\mathcal{H})$

$$\Pr(\cap A_i(R)) > 1 - K \cdot \frac{1}{K} = 0. \quad (103)$$

Thus, there must exist a precoding matrix \mathbf{P}_c for which the target $R < R_{\text{WC-CL}}(\mathcal{H})$ rate is achievable (via P-IF transmission) for the compound channel (89).

Figure 8 depicts the corresponding upper bounds on the gap-to-capacity for a $N_r \times 2$ channel for $K = 2, 3, 4$ users and for $C(\mathcal{H}) = 14$ bits. For calculating the upper bound on the guaranteed closed-loop gap, we use the tightest bound we have developed for $N_t = 2$ which is Corollary 1 of Lemma 4. We observe that

¹²IF-SIC is used for $N_t = 2$ since it provides improved bounds.

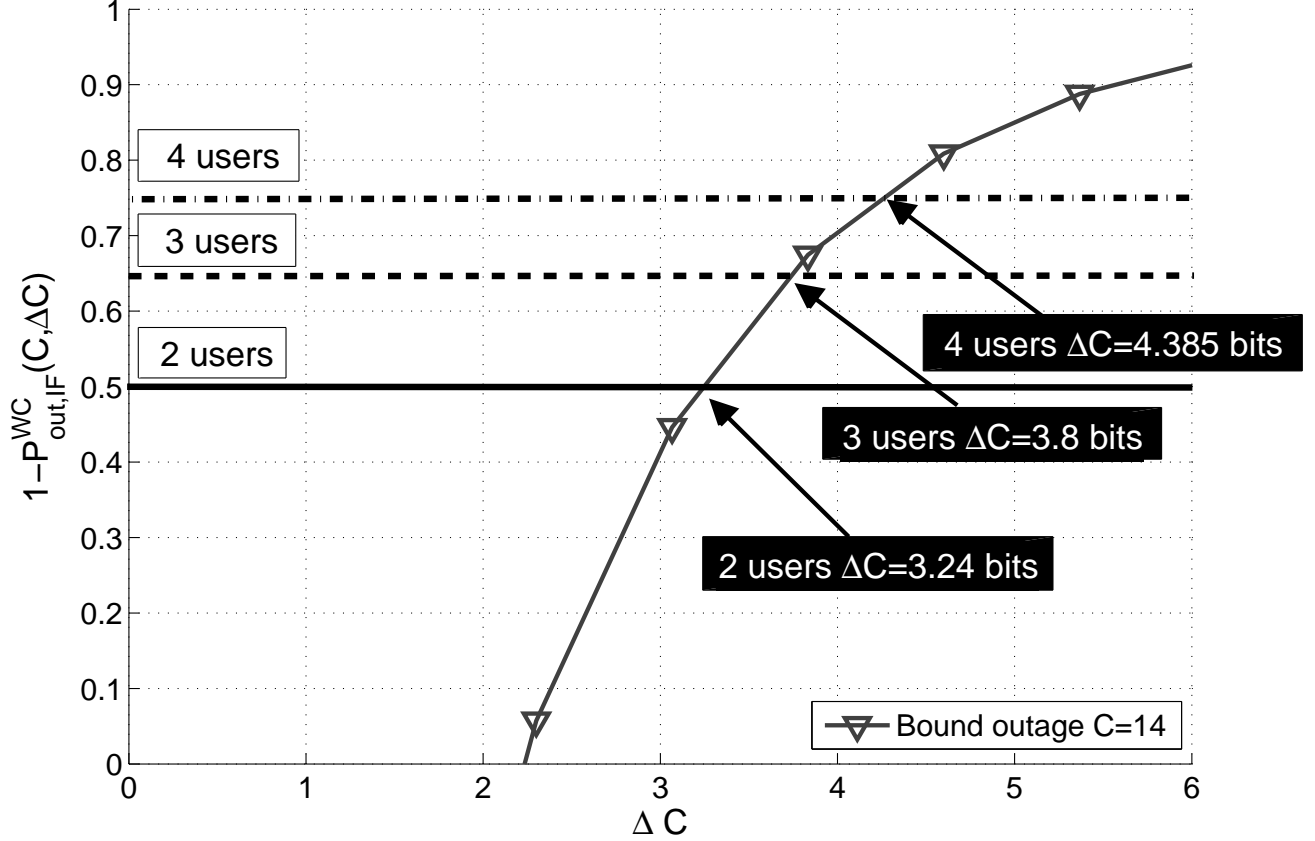


Fig. 8. Guaranteed closed performance for $N_r \times 2$ complex channels with 2,3,4 users.

- For 2 users, a rate of 10.76 bits is guaranteed (gap of 3.24 bits from capacity).
- For 3 users, a rate of 10.2 bits is guaranteed (gap of 3.8 bits from capacity).
- For 4 users, a rate of 9.615 bits is guaranteed (gap of 4.385 bits from capacity).

VI. CONCLUSION

We obtained analytical universal bounds for the outage probability of a transmission scheme employing randomly unitary precoding at the transmitter side and integer-forcing equalization at the receiver side. These bounds provide meaningful performance guarantees for transmission over MIMO channels that depend only on the channel's mutual information. Nonetheless, simulations suggest that there is still a considerable gap between the obtained bounds and the true (worst-case) outage probability of the examined scheme, calling for further work.

VII. ACKNOWLEDGEMENT

The authors are deeply grateful to Or Ordentlich whose work laid the foundation to the this paper and for many fruitful discussions.

APPENDIX A PROOF OF LEMMA 1

We start by expressing $\|\mathbf{D}^{1/2}\mathbf{V}\mathbf{a}\|$ equivalently as a complex expression. We note that \mathbf{a} (which is a vector of $2N_t$ real integers) can be viewed as a real representation of the complex vector \mathbf{a}_c such that

$$\mathbf{a} = \begin{bmatrix} \text{Re}(\mathbf{a}_c) \\ \text{Im}(\mathbf{a}_c) \end{bmatrix}. \quad (104)$$

Obviously, $\|\mathbf{a}\| = \|\mathbf{a}_c\|$.

Using this notation, and since \mathbf{D} and \mathbf{V} are real representation of complex matrices

$$\left\| \mathbf{D}^{1/2} \mathbf{V} \mathbf{a} \right\| = \left\| \mathbf{D}_c^{1/2} \mathbf{V}_c \mathbf{a}_c \right\|. \quad (105)$$

Now since \mathbf{V}_c is drawn from CUE, the distribution of $\left\| \mathbf{D}_c^{1/2} \mathbf{V}_c \mathbf{a}_c \right\|$ equals the distribution of $\left\| \mathbf{D}_c^{1/2} \mathbf{V}_c \begin{bmatrix} 1 & 0 & \dots & 0 \end{bmatrix}^T \|\mathbf{a}_c\| \right\|$. Note also that

$$\left\| \mathbf{D}_c^{1/2} \mathbf{V}_c \begin{bmatrix} 1 & 0 & \dots & 0 \end{bmatrix}^T \|\mathbf{a}_c\| \right\| = \quad (106)$$

$$\left\| \mathbf{D}_c^{1/2} \mathbf{V}_c \begin{bmatrix} 1 & 0 & \dots & 0 \end{bmatrix}^T \right\| \|\mathbf{a}\| = \quad (107)$$

$$\left\| \mathbf{D}_c^{1/2} \mathbf{v}_{c,1} \right\| \|\mathbf{a}\|. \quad (108)$$

where $\mathbf{v}_{c,1}$ is the first column of \mathbf{V}_c .

As described in [19], $\mathbf{v}_{c,1}$ is a vector uniformly distributed over the surface of the complex unit sphere. Such a vector can be generated by taking a vector with zero-mean i.i.d. complex Gaussian components and scaling it by its norm. The components of a vector uniformly distributed over the surface of the complex unit sphere can be written as

$$\mathbf{v}_{c,1} = \frac{G_i}{\sqrt{\sum_{i=1}^{N_t} |G_i|^2}}, \quad (109)$$

where G_i are zero-mean i.i.d. complex circularly symmetric Gaussian random variables.

This can be written over the reals as

$$\left\| \mathbf{D}^{1/2} \begin{bmatrix} \text{Re}(\mathbf{v}_{c,1}) \\ \text{Im}(\mathbf{v}_{c,1}) \end{bmatrix} \right\| \|\mathbf{a}\|. \quad (110)$$

Now, since the real and imaginary part of the complex Gaussian components are i.i.d. real Gaussian random variables, the resulting $2N_t \times 1$ vector

$$\mathbf{o} = \begin{bmatrix} \text{Re}(\mathbf{v}_{c,1}) \\ \text{Im}(\mathbf{v}_{c,1}) \end{bmatrix} \quad (111)$$

is uniformly distributed over the surface of a real unit ($2N_t$ -dimensional) sphere, hence

$$\left\| \mathbf{D}^{1/2} \mathbf{o} \right\| \|\mathbf{a}\|. \quad (112)$$

This vector can be interpreted as the first vector from a real matrix \mathbf{O} drawn from COE ensemble, and hence the distribution of (112) is the same as that of

$$\left\| \mathbf{D}^{1/2} \mathbf{O} \mathbf{a} \right\|. \quad (113)$$

It follows that

$$\Pr \left(\left\| \mathbf{D}^{1/2} \mathbf{V} \mathbf{a} \right\| < \sqrt{\beta} \right) = \Pr \left(\left\| \mathbf{D}^{1/2} \mathbf{O} \mathbf{a} \right\| < \sqrt{\beta} \right). \quad (114)$$

APPENDIX B PROOF OF THEOREM 1

From Lemma 2 we have

$$\Pr(R_{\text{IF}}(\mathbf{D}, \mathbf{V}) < C - \Delta C) \leq \quad (115)$$

$$\sum_{\mathbf{a} \in \mathbb{A}(\beta, d_{\min})} \frac{2N_t \left(2^{\frac{C-\Delta C}{N_t}} \alpha(N_t) \right)^{N_t-1/2}}{\|\mathbf{a}\|^{2N_t-1} 2^C \frac{2}{\sqrt{d_{\min}}}}. \quad (116)$$

Using Lemma 1 in [20] we have

$$\sum_{\mathbb{A}(\beta, d_{\min})} \frac{2N_t \sqrt{\beta}^{2N_t-1}}{\|\mathbf{a}\|^{2N_t-1} 2^C \frac{2}{\sqrt{d_{\min}}}} \leq \quad (117)$$

$$\sum_{k=1}^{\lfloor \sqrt{\frac{\beta}{d_{\min}}} \rfloor} \sum_{k \leq \|\mathbf{a}\| \leq k+1} \frac{2N_t \sqrt{\beta}^{2N_t-1}}{k^{2N_t-1} 2^C \frac{2}{\sqrt{d_{\min}}}}. \quad (118)$$

Denote $f = f(Nt, \beta, d_{\min}, C) = \frac{2N_t \sqrt{\beta}^{2N_t-1}}{2^C \frac{2}{\sqrt{d_{\min}}}}$. We have

$$\sum_{\mathbb{A}(\beta, d_{\min})} \frac{2N_t \sqrt{\beta}^{2N_t-1}}{\|\mathbf{a}\|^{2N_t-1} 2^C \frac{2}{\sqrt{d_{\min}}}} \leq \quad (119)$$

$$\sum_{k=1}^{\lfloor \sqrt{\frac{\beta}{d_{\min}}} \rfloor} \frac{f \Omega_{2N_t}}{k^{2N_t-1}} \left[\left(k + 1 + \frac{\sqrt{2N_t}}{2} \right)^{2N_t} - \left(k - \frac{\sqrt{2N_t}}{2} \right)^{2N_t} \right] \quad (120)$$

Using the binomial expansion, we have

$$\left[\left(k + 1 + \frac{\sqrt{2N_t}}{2} \right)^{2N_t} - \left(k - \frac{\sqrt{2N_t}}{2} \right)^{2N_t} \right] = \quad (121)$$

$$\sum_{i=1}^{2N_t} \binom{2N_t-i}{i} k^{2N_t-i} \left[\left(1 + \frac{\sqrt{2N_t}}{2} \right)^i - \left(-\frac{\sqrt{2N_t}}{2} \right)^i \right]. \quad (122)$$

We seek a constant c_1 such that (122) will be upper bounded by $c_1 k^{2N_t-1}$. That is, c_1 should satisfy

$$c_1 \geq \sum_{i=1}^{2N_t} \binom{2N_t-i}{i} k^{1-i} \left[\left(1 + \frac{\sqrt{2N_t}}{2} \right)^i - \left(-\frac{\sqrt{2N_t}}{2} \right)^i \right]. \quad (123)$$

It's easy to verify that the derivative (with respect to k) of the right side of (123) is negative. Thus its maximum is attained when $k = 1$, which means that

$$c_1 = \sum_{i=1}^{2N_t} \binom{2N_t-i}{i} \left[\left(1 + \frac{\sqrt{2N_t}}{2} \right)^i - \left(-\frac{\sqrt{2N_t}}{2} \right)^i \right] \quad (124)$$

$$= \left[\left(2 + \frac{\sqrt{2N_t}}{2} \right)^{2N_t} - \left(1 - \frac{\sqrt{2N_t}}{2} \right)^{2N_t} \right]. \quad (125)$$

Applying the bound (123) to (120), we get

$$\sum_{\mathbb{A}(\beta, d_{\min})} \frac{2N_t \sqrt{\beta}^{2N_t-1}}{\|\mathbf{a}\|^{2N_t-1} 2^C \frac{2}{\sqrt{d_{\min}}}} \leq \sum_{k=1}^{\lfloor \sqrt{\frac{\beta}{d_{\min}}} \rfloor} f \Omega_{2N_t} \cdot c_1 \quad (126)$$

$$= \sum_{k=1}^{\lfloor \sqrt{\frac{\beta}{d_{\min}}} \rfloor} \frac{2c_1 N_t \sqrt{\beta}^{2N_t-1}}{2^C \frac{2}{\sqrt{d_{\min}}}} \Omega_{2N_t} \quad (127)$$

$$\leq \int_{k=0}^{\sqrt{\frac{\beta}{d_{\min}}}} \frac{2c_1 N_t \sqrt{\beta}^{2N_t-1}}{2^C \frac{2}{\sqrt{d_{\min}}}} \Omega_{2N_t} dk \quad (128)$$

$$= \frac{2c_1 N_t \sqrt{\beta}^{2N_t-1}}{2^C \frac{2}{\sqrt{d_{\min}}}} \sqrt{\frac{\beta}{d_{\min}}} \Omega_{2N_t} \quad (129)$$

$$= \frac{c_1 N_t \beta^{N_t}}{2^C} \Omega_{2N_t}. \quad (130)$$

Now since $\beta = 2^{\frac{C-\Delta C}{N_t}} (2N_t)^2 \alpha(N_t)$ we have

$$\sum_{\mathbb{A}(\beta, d_{\min})} \Pr \left(\|\mathbf{D}^{1/2} \mathbf{o}_{\|\mathbf{a}\|}\| < \sqrt{\beta} \right) \quad (131)$$

$$\leq \frac{c_1 N_t \left(2^{\frac{C-\Delta C}{N_t}} \alpha(N_t) \right)^{N_t}}{2^C} \Omega_{2N_t} \quad (132)$$

$$= \frac{c_1 N_t 2^{C-\Delta C} \alpha(N_t)^{N_t}}{2^C} \Omega_{2N_t} \quad (133)$$

$$= c_1 N_t \alpha(N_t)^{N_t} \frac{\pi^{N_t}}{\Gamma(N_t + 1)} 2^{-\Delta C}. \quad (134)$$

Hence,

$$\sum_{\mathbb{A}(\beta, d_{\min})} \frac{2N_t \sqrt{\beta}^{2N_t-1}}{\|\mathbf{a}\|^{2N_t-1} 2^C \frac{2}{\sqrt{d_{\min}}}} \Omega_{2N_t} \leq c(N_t) 2^{-\Delta C} \quad (135)$$

where

$$c(N_t) = \left[\left(2 + \frac{\sqrt{2N_t}}{2} \right)^{2N_t} - \left(1 - \frac{\sqrt{2N_t}}{2} \right)^{2N_t} \right] N_t \alpha(N_t)^{N_t} \frac{\pi^{N_t}}{\Gamma(N_t + 1)} \quad (136)$$

is a constant that depends only on N_t . We note that (136) does not depend on \mathbf{D} hence it holds also for the supremum over \mathbf{D} . Recalling (28), we have

$$P_{\text{out,IF}}^{\text{WC}}(C, \Delta C) = \sup_{\mathbf{D} \in \mathbb{D}(C)} \Pr(R_{\text{IF}}(\mathbf{D}, \mathbf{V}) < C - \Delta C) \quad (137)$$

$$= \sup_{\mathbf{D} \in \mathbb{D}(C)} \sum_{\mathbb{A}(\beta, d_{\min})} \frac{2N_t \sqrt{\beta}^{2N_t-1}}{\|\mathbf{a}\|^{2N_t-1} 2^C \frac{2}{\sqrt{d_{\min}}}} \Omega_{2N_t} \quad (138)$$

$$\leq c(N_t) 2^{-\Delta C}. \quad (139)$$

APPENDIX C

TIGHTER BOUNDS FOR $N_r \times 2$ CHANNELS

We start by analyzing $N_r \times 2$ complex channels with IF-SIC. As we mentioned in Section III-C, when using complex precoding, the rates appear in pairs. Denoting

$$R^{1,\text{IF}}(\mathbf{D}, \mathbf{V}) = R_{1,\text{IF}}(\mathbf{D}, \mathbf{V}) = R_{2,\text{IF}}(\mathbf{D}, \mathbf{V}) \quad (140)$$

$$R^{2,\text{IF}}(\mathbf{D}, \mathbf{V}) = R_{3,\text{IF}}(\mathbf{D}, \mathbf{V}) = R_{4,\text{IF}}(\mathbf{D}, \mathbf{V}) \quad (141)$$

$$R^{1,\text{IF-SIC}}(\mathbf{D}, \mathbf{V}) = R_{1,\text{IF-SIC}}(\mathbf{D}, \mathbf{V}) = R_{2,\text{IF-SIC}}(\mathbf{D}, \mathbf{V}) \quad (142)$$

$$R^{2,\text{IF-SIC}}(\mathbf{D}, \mathbf{V}) = R_{3,\text{IF-SIC}}(\mathbf{D}, \mathbf{V}) = R_{4,\text{IF-SIC}}(\mathbf{D}, \mathbf{V}), \quad (143)$$

we have

$$C = 2(R^{1,\text{IF-SIC}}(\mathbf{D}, \mathbf{V}) + R^{2,\text{IF-SIC}}(\mathbf{D}, \mathbf{V})). \quad (144)$$

Since we use IF-SIC with equal rate per stream, we have the following

$$R_{\text{IF-SIC}}(\mathbf{D}, \mathbf{V}) = 4 \min(R^{1,\text{IF-SIC}}(\mathbf{D}, \mathbf{V}), R^{2,\text{IF-SIC}}(\mathbf{D}, \mathbf{V})). \quad (145)$$

When applying SIC, one decodes first the equation with the *highest* SNR (highest rate). Since for this equation SIC has no effect it follows that

$$R^{1,\text{IF-SIC}}(\mathbf{D}, \mathbf{V}) = R^{1,\text{IF}}(\mathbf{D}, \mathbf{V}) \quad (146)$$

Substituting (146) into (144), we have

$$R^{2,\text{IF-SIC}}(\mathbf{D}, \mathbf{V}) = \frac{C}{2} - R^{1,\text{IF}}(\mathbf{D}, \mathbf{V}) \quad (147)$$

Now, from [21] Theorem 3 (with equivalent four real dimensions), we have

$$2(R^{1,\text{IF}}(\mathbf{D}, \mathbf{V}) + R^{2,\text{IF}}(\mathbf{D}, \mathbf{V})) > C - 4 \quad (148)$$

$$R^{1,\text{IF}}(\mathbf{D}, \mathbf{V}) + R^{2,\text{IF}}(\mathbf{D}, \mathbf{V}) > \frac{C - 4}{2}. \quad (149)$$

Since $R^{1,\text{IF}}(\mathbf{D}, \mathbf{V}) > R^{2,\text{IF}}(\mathbf{D}, \mathbf{V})$, it follows that

$$R^{1,\text{IF}}(\mathbf{D}, \mathbf{V}) > \frac{C - 4}{4}. \quad (150)$$

We conclude that

$$R_{\text{IF-SIC}}(\mathbf{D}, \mathbf{V}) = 4 \min(R^{1,\text{IF-SIC}}(\mathbf{D}, \mathbf{V}), R^{2,\text{IF-SIC}}(\mathbf{D}, \mathbf{V})) \quad (151)$$

$$= 4 \min\left(R^{1,\text{IF}}(\mathbf{D}, \mathbf{V}), \frac{C}{2} - R^{1,\text{IF}}(\mathbf{D}, \mathbf{V})\right) \quad (152)$$

$$> 4 \min\left(\frac{C - 4}{4}, \frac{C}{2} - R^{1,\text{IF}}(\mathbf{D}, \mathbf{V})\right) \quad (153)$$

$$= \min(C - 1, 2C - 4R^{1,\text{IF}}(\mathbf{D}, \mathbf{V})). \quad (154)$$

From this point we analyze the outage for $C > 1$ and target rates which are lower than $C - 1$, so that the inequality $2C - 4R^{1,\text{IF}}(\mathbf{D}, \mathbf{V}) < C - 1$ is satisfied. Thus, we consider gap-to-capacity such that $\Delta C > 1$. Our goal is to bound

$$\Pr(R_{\text{IF-SIC}}(\mathbf{D}, \mathbf{V}) < C - \Delta C) = \quad (155)$$

$$\Pr\left(2C - 4\frac{1}{2}\log\left(\frac{1}{\lambda_1^2(\Lambda)}\right) < C - \Delta C\right) = \quad (156)$$

$$\Pr\left(-2\log\left(\frac{1}{\lambda_1^2(\Lambda)}\right) < -(C + \Delta C)\right) = \quad (157)$$

$$\Pr\left(\lambda_1^2(\Lambda) < 2^{-1/2(C+\Delta C)}\right). \quad (158)$$

We are ready to prove Lemma 4 and Theorem 2.

Proof: Recalling (83), we begin by bounding $\Pr(R_{\text{IF-SIC}}(\mathbf{D}, \mathbf{V}) < C - \Delta C)$.

For some $\beta > 0$, let us upper bound the probability

$$\Pr(\lambda_1^2(\Lambda) < \beta) = \Pr(\lambda_1(\Lambda) < \sqrt{\beta}) \quad (159)$$

with respect to matrix \mathbf{D} . Note that the event $\lambda_1(\Lambda) < \sqrt{\beta}$ is equivalent to the event

$$\bigcup_{\mathbf{a} \in \mathbb{Z}^{2N_t} \setminus \{0\}} \|\mathbf{D}^{1/2} \mathbf{V} \mathbf{a}\| < \sqrt{\beta}. \quad (160)$$

Applying the union bound yields

$$\Pr(\lambda_1(\Lambda) < \sqrt{\beta}) < \sum_{\mathbf{a} \in \mathbb{Z}^{2N_t}} \Pr(\|\mathbf{D}^{1/2} \mathbf{V} \mathbf{a}\| < \sqrt{\beta}). \quad (161)$$

We note that if $\frac{\|\mathbf{a}\|}{d_{\max}} > \sqrt{\beta}$, we have

$$\Pr(\|\mathbf{D}^{1/2} \mathbf{V} \mathbf{a}\| < \sqrt{\beta}) = 0. \quad (162)$$

We therefore define the set of relevant vectors \mathbf{a} as

$$\mathbb{A}(\beta, d_{\min}) = \left\{ \mathbf{a} : 0 < \|\mathbf{a}\| < \sqrt{\beta d_{\max}} \right\}. \quad (163)$$

Applying the union bound, we get

$$\Pr(\lambda_1(\Lambda^*) < \sqrt{\beta}) < \sum_{\mathbf{a} \in \mathbb{A}(\beta, d_{\min})} \Pr(\|\mathbf{D}^{1/2} \mathbf{V}^T \mathbf{a}\| < \sqrt{\beta}). \quad (164)$$

We now apply a similar derivation to the general case considered in Section IV. Applying Lemma 1, we have

$$\Pr(\|\mathbf{D}^{1/2} \mathbf{V}^T \mathbf{a}\| < \sqrt{\beta}) = \Pr(\|\mathbf{D}^{1/2} \mathbf{O}^T \mathbf{a}\| < \sqrt{\beta}). \quad (165)$$

where \mathbf{O} is drawn from the COE. Hence, we can apply the same geometric interpretation as in Section IV and interpret

$\Pr(\|\mathbf{D}^{1/2}\mathbf{O}^T\mathbf{a}\| < \sqrt{\beta})$ as the ratio of the surface area of the four-dimensional ellipsoid inside a ball with radius $\sqrt{\beta}$ and the surface area of this ellipsoid. The axes of this ellipsoid are defined as

$$x_i = \frac{\|\mathbf{a}\|}{\sqrt{d_i}}. \quad (166)$$

For the case of 4 real dimensions (56) can be written as

$$\sum_{\mathbb{A}(\beta, d_{\min})} \Pr\left(\|\mathbf{D}^{1/2}\mathbf{O}\|_{\mathbf{a}} < \sqrt{\beta}\right) = \sum_{\mathbb{A}(\beta, d_{\min})} \frac{\text{CAP}_{\text{ell}}}{L(x_1, x_2, x_3, x_4)} \quad (167)$$

where

$$\text{CAP}_{\text{ell}} < A_4(\sqrt{\beta}) = 4\frac{\pi^2}{2}\sqrt{\beta}^3 \triangleq \overline{\text{CAP}_{\text{ell}}}, \quad (168)$$

and

$$L(x_1, x_2, x_3, x_4) > \frac{\pi^2}{2} \frac{\|\mathbf{a}\|^4}{\prod_{i=1}^4 \sqrt{d_i}} \left(\frac{2\sqrt{d_{\min}}}{\|\mathbf{a}\|} + \frac{2\sqrt{d_{\max}}}{\|\mathbf{a}\|} \right) \quad (169)$$

$$> \pi^2 \frac{\|\mathbf{a}\|^3}{2^C} \left(\sqrt{d_{\max}} \right) \triangleq \underline{L}(x_1, x_2, x_3, x_4). \quad (170)$$

Substituting (168) and (170) in (167), we obtain

$$\Pr(R_{\text{IF-SIC}}(\mathbf{D}, \mathbf{V}) < C - \Delta C) < \sum_{\mathbb{A}(\beta, d_{\min})} \frac{2\pi^2\sqrt{\beta}^3}{\pi^2 \frac{\|\mathbf{a}\|^3}{2^C} (\sqrt{d_{\max}})}. \quad (171)$$

Recalling that $\beta = 2^{1/2(C+\Delta C)}$, we get that for $\Delta C < 1$

$$\Pr(R_{\text{IF-SIC}}(\mathbf{D}, \mathbf{V}) < C - \Delta C) < \sum_{\mathbb{A}(\beta, d_{\min})} \frac{2\pi^2 2^{3/4(C+\Delta C)}}{\pi^2 \frac{\|\mathbf{a}\|^3}{2^C} (\sqrt{d_{\max}})}, \quad (172)$$

which proves Lemma 4.

To establish Theorem 2 we follow the footsteps of the proof of Theorem 1, to obtain

$$\sum_{\mathbb{A}(\beta, d_{\min})} \Pr\left(\|\mathbf{D}^{1/2}\mathbf{O}\|_{\mathbf{a}} < \sqrt{\beta}\right) < \sum_{\mathbb{A}(\beta, d_{\min})} \frac{2\pi^2\sqrt{\beta}^3}{\pi^2 \frac{\|\mathbf{a}\|^3}{2^C} (\sqrt{d_{\max}})} \quad (173)$$

$$= \sum_{\mathbb{A}(\beta, d_{\min})} \frac{2\sqrt{\beta}^3 2^C}{\|\mathbf{a}\|^3 (\sqrt{d_{\max}})} \quad (174)$$

$$< \sum_{k=1}^{\lfloor \sqrt{\beta d_{\max}} \rfloor} \sum_{k \leq \|\mathbf{a}\| \leq k+1} \frac{2\sqrt{\beta}^3 2^C}{k^3 (\sqrt{d_{\max}})} \quad (175)$$

$$< \sum_{k=1}^{\lfloor \sqrt{\beta d_{\max}} \rfloor} \frac{2\sqrt{\beta}^3 2^C}{k^3 (\sqrt{d_{\max}})} \frac{\pi^2}{2} \left[(k+2)^4 - (k-1)^4 \right] \quad (176)$$

$$< \sum_{k=1}^{\lfloor \sqrt{\beta d_{\max}} \rfloor} \frac{\sqrt{\beta}^3 2^C \pi^2}{k^3 (\sqrt{d_{\max}})} 81k^3 \quad (177)$$

$$< \sum_{k=1}^{\lfloor \sqrt{\beta d_{\max}} \rfloor} \frac{\sqrt{\beta}^3 2^C \pi^2 81}{(\sqrt{d_{\max}})} \quad (178)$$

$$< \int_0^{\sqrt{\beta d_{\max}}} \frac{\sqrt{\beta}^3 2^C \pi^2 81}{(\sqrt{d_{\max}})} dk \quad (179)$$

$$= \frac{\sqrt{\beta^3} 2^C \pi^2 81 \sqrt{\beta d_{\max}}}{(\sqrt{d_{\max}})} \quad (180)$$

$$= \beta^2 2^C \pi^2 81. \quad (181)$$

This does not depend on \mathbf{D} hence it holds also for the supremum over \mathbf{D} . Now since $\beta = 2^{-1/2(C+\Delta C)}$ we get

$$P_{\text{out,IF-SIC}}^{\text{WC}}(C, \Delta C) \leq 81 \cdot \pi^2 \cdot 2^C 2^{-C-\Delta C} \quad (182)$$

$$= 81 \cdot \pi^2 \cdot 2^{-\Delta C}. \quad (183)$$

■

APPENDIX D

DMT CURVE OBTAINED USING THEOREM 1

We analyze the diversity-multiplexing tradeoff for i.i.d. Rayleigh fading. Hence, we assume that the entries of the matrix $\mathbf{H}_c \in \mathbb{C}^{N_r \times N_t}$ are statistically independent and each is a unit variance circularly symmetric complex Gaussian random variable.

To obtain the tradeoff curve, we are interested in

$$\Pr \left(R_{\text{IF}}(\sqrt{\text{SNR}} \cdot \mathbf{H}_c) < r \log \text{SNR} \right) \quad (184)$$

$$= \int_{c=0}^{\infty} f_{\text{Ray}}(c, \text{SNR}) \cdot \Pr \left(R_{\text{IF}}(\sqrt{\text{SNR}} \cdot \mathbf{H}_c) < r \log \text{SNR} | C_{\text{WI}} = c \right) dc \quad (185)$$

where we denote $F_{\text{Ray}}(c, \text{SNR}) = \Pr(\log(\mathbf{I} + \text{SNR} \mathbf{H}_c \mathbf{H}_c^H) \leq c)$ and $f_{\text{Ray}}(c, \text{SNR})$ as its derivative with respect to c . Recalling (6) and substituting (70) of Theorem 1, we have

$$\Pr(R_{\text{IF}}(\sqrt{\text{SNR}} \cdot \mathbf{H}_c) < r \log \text{SNR} | C_{\text{WI}} = c) \leq P_{\text{out,IF}}^{\text{WC}}(c, r \log \text{SNR}) \quad (186)$$

$$\leq c_1 2^{-(c-r \log \text{SNR})}. \quad (187)$$

Substituting (187) into (185), we get

$$P \left(R_{\text{IF}}(\sqrt{\text{SNR}} \cdot \mathbf{H}_c) r \log \text{SNR} \right) \leq \int_{c=0}^{\infty} f_{\text{Ray}}(c, \text{SNR}) c_1 2^{-(c-r \log \text{SNR})} dc \quad (188)$$

$$= c_1 \log \text{SNR} \int_{\rho=0}^{\infty} f_{\text{Ray}}(\rho \log \text{SNR}) 2^{-(\rho-r) \log \text{SNR}} d\rho \quad (189)$$

$$\doteq \int_{\rho=0}^{\infty} \frac{d}{d\rho} \Pr(C_{\text{Ray}} < \rho \log \text{SNR}) \text{SNR}^{-(\rho-r)} d\rho \quad (190)$$

$$\doteq \int_{\rho=0}^{\infty} \text{SNR}^{-(\rho-r)} \frac{d}{d\rho} (\text{SNR}^{-d^*(\rho)}) d\rho, \quad (191)$$

where $d^*(\rho)$ is the optimal DMT. Now, this integral can equivalently written as

$$\sum_{k=0}^{\min(N_t, N_r)-1} \int_{\rho=k}^{k+1} \text{SNR}^{-(\rho-r)} \frac{d}{d\rho} \text{SNR}^{-d^*(\rho)} d\rho. \quad (192)$$

Since $d^*(\rho)$ is piecewise linear within each interval $[k, k+1]$, we have

$$\frac{d}{d\rho} \text{SNR}^{-d^*(\rho)} = \log(\text{SNR}) \cdot \text{SNR}^{-d^*(\rho)} \cdot \frac{d}{d\rho} -d^*(\rho) \quad (193)$$

$$\doteq \text{SNR}^{-d^*(\rho)}. \quad (194)$$

Hence,

$$P \left(R_{\text{IF}}(\sqrt{\text{SNR}} \cdot \mathbf{H}_c) \right) \doteq \sum_{k=0}^{\min(N_t, N_r)-1} \int_{\rho=k}^{k+1} \text{SNR}^{r-\rho-d^*(\rho)} d\rho \quad (195)$$

$$\doteq \text{SNR}^{-d_{\text{Thm1}}(r)}, \quad (196)$$

where $d^*(\rho)$ is the optimal DMT and

$$d_{\text{Thm1}}(r) = \min_{k=0, \dots, \min(N_t, N_r)-1} \min_{k \leq \rho \leq k+1} \{d^*(\rho) + \rho - r\}. \quad (197)$$

By Theorem 2 in [2], the optimal trade-off curve is given by the piecewise-linear function connecting the points $(k, d^*(k))$, $k = 0, 1, \dots, \min(N_t, N_r)$, where $d^*(k) = (N_t - k)(N_r - k)$. It follows that

$$\min_{k \leq \rho \leq k+1} \{d^*(\rho) + \rho\} = d^*(k+1) + k + 1, \quad (198)$$

and

$$\min_{k=0, \dots, \min(N_t, N_r)-1} \{d^*(k+1) + k + 1\} = \min(N_t, N_r). \quad (199)$$

Hence, the DMT that is achievable using Theorem 1 is

$$d_{\text{Thm1}}(r) = \min(N_t, N_r) - r, \quad (200)$$

which is the optimal receive diversity only when $N_t = N_r$.

REFERENCES

- [1] I. E. Telatar, "Capacity of multi-antenna Gaussian channels," *Eur. Trans. Telecommun.*, vol. 10, pp. 585–598, Nov 1999.
- [2] D. Tse and P. Viswanath, *Fundamentals of Wireless Communication*. Cambridge university press, 2005.
- [3] B. Hochwald and S. ten Brink, "Achieving near-capacity on a multiple-antenna channel," *Communications, IEEE Transactions on*, vol. 51, no. 3, pp. 389–399, March 2003.
- [4] O. Ordentlich and U. Erez, "Precoded integer-forcing universally achieves the MIMO capacity to within a constant gap," *IEEE Transactions on Information Theory*, vol. 61, no. 1, pp. 323–340, Jan 2015.
- [5] P. Elia, K. Kumar, S. Pawar, P. Kumar, and H. feng Lu, "Space-time codes meeting the diversity-multiplexing gain tradeoff with low signalling complexity," *Information Theory, IEEE Transactions on*, vol. 52, no. 9, pp. 3869–3884, Sept 2006.
- [6] E. Larsson, "Constellation randomization (CoRa) for outage performance improvement on MIMO channels," in *Global Telecommunications Conference, 2004. GLOBECOM '04. IEEE*, vol. 1, Nov 2004, pp. 386–390 Vol.1.
- [7] Y. Li, C. Georgiades, and G. Huang, "Transmit diversity over quasi-static fading channels using multiple antennas and random signal mapping," *Communications, IEEE Transactions on*, vol. 51, no. 11, pp. 1918–1926, Nov 2003.
- [8] J. Zhan, B. Nazer, U. Erez, and M. Gastpar, "Integer-forcing linear receivers," vol. 60, no. 12, Dec 2014, pp. 7661–7685.
- [9] O. Ordentlich, U. Erez, and B. Nazer, "Successive integer-forcing and its sum-rate optimality," pp. 282–292, Oct 2013.
- [10] M. Mehta, *Random matrices and the statistical theory of energy level*. Academic Press, 1967.
- [11] J. C. Lagarias, H. W. Lenstra Jr, and C.-P. Schnorr, "Korkin-zolotarev bases and successive minima of a lattice and its reciprocal lattice," *Combinatorica*, vol. 10, no. 4, pp. 333–348, 1990.
- [12] W. Banaszczyk, "New bounds in some transference theorems in the geometry of numbers," *Mathematische Annalen*, vol. 296, no. 1, pp. 625–635, 1993.
- [13] H. F. Blichfeldt, "The minimum value of quadratic forms, and the closest packing of spheres," *Mathematische Annalen*, vol. 101, no. 1, pp. 605–608, 1929.
- [14] G. Tee, "Surface area and capacity of ellipsoids in n dimensions.(CITR)," CITR, The University of Auckland, New Zealand, Tech. Rep., 2004.
- [15] A. Sakzad, J. Harshan, and E. Viterbo, "On complex LLL algorithm for integer forcing linear receivers," in *Communications Theory Workshop (AusCTW), 2013 Australian*, 2013, pp. 13–17.
- [16] R. Fischer, J. N. Laneman, and S. Stern, "Factorization approaches in lattice-reduction-aided and integer-forcing equalization," in *Proc. Int. Zurich Seminar on Communications*, Mar. 2012.
- [17] L. Zheng and D. Tse, "Diversity and multiplexing: a fundamental tradeoff in multiple-antenna channels," *Information Theory, IEEE Transactions on*, vol. 49, no. 5, pp. 1073–1096, May 2003.
- [18] A. Khina, I. Livni, A. Hitron, and U. Erez, "Joint unitary triangularization for Gaussian multi-user MIMO networks," *IEEE Transactions on Information Theory*, vol. 61, no. 5, pp. 2662–2692, May 2015.
- [19] A. Narula, M. Trott, and G. W. Wornell, "Performance limits of coded diversity methods for transmitter antenna arrays," *Information Theory, IEEE Transactions on*, vol. 45, no. 7, pp. 2418–2433, Nov 1999.
- [20] O. Ordentlich and U. Erez, "A simple proof for the existence of good pairs of nested lattices," in *Electrical Electronics Engineers in Israel (IEEEI), 2012 IEEE 27th Convention of*, Nov 2012, pp. 1–12.
- [21] O. Ordentlich, U. Erez, and B. Nazer, "The approximate sum capacity of the symmetric Gaussian-user interference channel," *Information Theory, IEEE Transactions on*, vol. 60, no. 6, pp. 3450–3482, June 2014.